#11

APPENDIX A

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:

Kevin P. Baker et al.

Serial No. 09/944,396

Filing Date: August 30, 2001

For SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

Examiner: Kemmerer, E.

Group Art Unit No.: 1646

DECLARATION OF AUDREY D. GODDARD, Ph.D UNDER 37 C.F.R. § 1.132

Assistant Commissioner of Patents
Washington, D.C. 20231

Sir:

I, Audrey D. Goddard, Ph.D. do hereby declare and say as follows:

1. I am a Senior Clinical Scientist at the Experimental Medicine/BioOncology, Medical Affairs Department of Genentech, Inc., South San Francisco, California 94080.

2. Between 1993 and 2001, I headed the DNA Sequencing Laboratory at the Molecular Biology Department of Genentech, Inc. During this time, my responsibilities included the identification and characterization of genes contributing to the oncogenic process, and determination of the chromosomal localization of novel genes.

3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

4. I am familiar with a variety of techniques known in the art for detecting and quantifying the amplification of oncogenes in cancer, including the quantitative TaqMan PCR (i.e., "gene amplification") assay described in the above captioned patent application.

5. The TaqMan PCR assay is described, for example, in the following scientific publications: Higuchi *et al.*, Biotechnology 10:413-417 (1992) (Exhibit B); Livak *et al.*, PCR Methods Appl., 4:357-362 (1995) (Exhibit C) and Heid *et al.*, Genome Res. 6:986-994 (1996) (Exhibit D). Briefly, the assay is based on the principle that successful PCR yields a fluorescent signal due to Taq DNA polymerase-mediated exonuclease digestion of a fluorescently labeled oligonucleotide that is homologous to a sequence between two PCR primers. The extent of digestion depends directly on the amount of PCR, and can be quantified accurately by measuring the increment in fluorescence that results from decreased energy transfer. This is an extremely sensitive technique, which allows detection in the exponential phase of the PCR reaction and, as a result, leads to accurate determination of gene copy number.

6. The quantitative fluorescent TaqMan PCR assay has been extensively and successfully used to characterize genes involved in cancer development and progression. Amplification of protooncogenes has been studied in a variety of human tumors, and is widely considered as having etiological, diagnostic and prognostic significance. This use of the quantitative TaqMan PCR assay is exemplified by the following scientific publications: Pennica *et al.*, Proc. Natl. Acad. Sci. USA 95(25):14717-14722 (1998) (Exhibit E); Pitti *et al.*, Nature 396(6712):699-703 (1998) (Exhibit F) and Bieche *et al.*, Int. J. Cancer 78:661-666 (1998) (Exhibit G), the first two of which I am co-author. In particular, Pennica *et al.* have used the quantitative TaqMan PCR assay to study relative gene amplification of WISP and c-myc in various cell lines, colorectal tumors and normal mucosa. Pitti *et al.* studied the genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer, using the quantitative TaqMan PCR assay. Bieche *et al.* used the assay to study gene amplification in breast cancer.

7. It is my personal experience that the quantitative TaqMan PCR technique is technically sensitive enough to detect at least a 2-fold increase in gene copy number relative to control. It is further my considered scientific opinion that an at least 2-fold increase in gene copy number in a tumor tissue sample relative to a normal (i.e., non-tumor) sample is significant and useful in that the detected increase in gene copy number in the tumor sample relative to the normal sample serves as a basis for using relative gene copy number as quantitated by the TaqMan PCR technique as a diagnostic marker for the presence or absence of tumor in a tissue sample of unknown pathology. Accordingly, a gene identified as being amplified at least 2-fold by the quantitative TaqMan PCR assay in a tumor sample relative to a normal sample is useful as a marker for the diagnosis of cancer, for monitoring cancer development and/or for measuring the efficacy of cancer therapy.

8. I declare further that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true. I declare that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Jan. 16, 2003
Date

Audrey Goddard
Audrey D. Goddard, Ph.D.

AUDREY D. GODDARD, Ph.D.

Genentech, Inc.
1 DNA Way
South San Francisco, CA, 94080
650.225.6429
goddarda@gene.com

110 Congo St.
San Francisco, CA, 94131
415.841.9154
415.819.2247 (mobile)
agoddard@pacbell.net

PROFESSIONAL EXPERIENCE

Genentech, Inc. **1993-present**
South San Francisco, CA

2001 - present **Senior Clinical Scientist**
Experimental Medicine / BioOncology, Medical Affairs

Responsibilities:
- *Companion diagnostic oncology products*
- *Acquisition of clinical samples from Genentech's clinical trials for translational research*
- *Translational research using clinical specimen and data for drug development and diagnostics*
- *Member of Development Science Review Committee, Diagnostic Oversight Team, 21 CFR Part 11 Subteam*

Interests:
- *Ethical and legal implications of experiments with clinical specimens and data*
- *Application of pharmacogenomics in clinical trials*

1998 - 2001 **Senior Scientist**
Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

Responsibilities:
- *Management of a laboratory of up to nineteen –including postdoctoral fellow, associate scientist, senior research associate and research assistants/associate levels*
- *Management of a $750K budget*
- *DNA sequencing core facility supporting a 350+ person research facility.*
- *DNA sequencing for high throughput gene discovery, - ESTs, cDNAs, and constructs*
- *Genomic sequence analysis and gene identification*
- *DNA sequence and primary protein analysis*

Research:
- *Chromosomal localization of novel genes*
- *Identification and characterization of genes contributing to the oncogenic process*
- *Identification and characterization of genes contributing to inflammatory diseases*
- *Design and development of schemes for high throughput genomic DNA sequence analysis*
- *Candidate gene prediction and evaluation*

1993 - 1998 Scientist
Head of the DNA Sequencing Laboratory, Molecular Biology Department, Research

Responsibilities
- *DNA sequencing core facility supporting a 350+ person research facility*
- *Assumed responsibility for a pre-existing team of five technicians and expanded the group into fifteen, introducing a level of middle management and additional areas of research*
- *Participated in the development of the basic plan for high throughput secreted protein discovery program – sequencing strategies, data analysis and tracking, database design*
- *High throughput EST and cDNA sequencing for new gene identification.*
- *Design and implementation of analysis tools required for high throughput gene identification.*
- *Chromosomal localization of genes encoding novel secreted proteins.*

Research:
- *Genomic sequence scanning for new gene discovery.*
- *Development of signal peptide selection methods.*
- *Evaluation of candidate disease genes.*
- *Growth hormone receptor gene SNPs in children with idiopathic short stature*

Imperial Cancer Research Fund **1989-1992**
London, UK with Dr. Ellen Solomon

6/89 –12/92 Postdoctoral Fellow
- Cloning and characterization of the genes fused at the acute promyelocytic leukemia translocation breakpoints on chromosomes 17 and 15.
- Prepared a successfully funded European Union multi-center grant application

McMaster University **1983**
Hamilton, Ontario, Canada with Dr. G. D. Sweeney

5/83 – 8/83: NSERC Summer Student
- *In vitro* metabolism of β-naphthoflavone in C57Bl/6J and DBA mice

EDUCATION

Ph.D. "Phenotypic and genotypic effects of mutations in the human retinoblastoma gene." **Supervisor:** Dr. R. A. Phillips	University of Toronto Toronto, Ontario, Canada. Department of Medical Biophysics.	1989
Honours B.Sc "The *In vitro* metabolism of the cytochrome P-448 inducer β-naphthoflavone in C57BL/6J mice." **Supervisor:** Dr. G. D. Sweeney	McMaster University, Hamilton, Ontario, Canada. Department of Biochemistry	1983

ACADEMIC AWARDS

Imperial Cancer Research Fund Postdoctoral Fellowship	1989-1992
Medical Research Council Studentship	1983-1988
NSERC Undergraduate Summer Research Award	1983
Society of Chemical Industry Merit Award (Hons. Biochem.)	1983
Dr. Harry Lyman Hooker Scholarship	1981-1983
J.L.W. Gill Scholarship	1981-1982
Business and Professional Women's Club Scholarship	1980-1981
Wyerhauser Foundation Scholarship	1979-1980

INVITED PRESENTATIONS

Genentech's gene discovery pipeline: High throughput identification, cloning and characterization of novel genes. Functional Genomics: From Genome to Function. Litchfield Park, AZ. USA. October 2000

High throughput identification, cloning and characterization of novel genes. G2K:Back to Science, Advances in Genome Biology and Technology I. Marco Island, FL, USA. February 2000

Quality control in DNA Sequencing: The use of Phred and Phrap. Bay Area Sequencing Users Meeting, Berkeley, CA, USA. April 1999

High throughput secreted protein identification and cloning. Tenth International Genome Sequencing and Analysis Conference, Miami, FL, USA. September 1998

The evolution of DNA sequencing: The Genentech perspective. Bay Area Sequencing Users Meeting, Berkeley, CA. USA. May 1998

Partial Growth Hormone Insensitivity: The role of GH-receptor mutations in Idiopathic Short Stature. Tenth Annual National Cooperative Growth Study Investigators Meeting, San Francisco, CA, USA. October, 1996

Growth hormone (GH) receptor defects are present in selected children with non-GH-deficient short stature: A molecular basis for partial GH-insensitivity. 76th Annual Meeting of The Endocrine Society, Anaheim, CA, USA. June 1994

A previously uncharacterized gene, myl, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. XV International Association for Comparative Research on Leukemia and Related Disease, Padua, Italy. October 1991

PATENTS

Goddard A, Godowski PJ, Gurney AL. NL2 Tie ligand homologue polypeptide. Patent Number: 6,455,496. Date of Patent: Sept. 24, 2002.

Goddard A, Godowski PJ and Gurney AL. NL3 Tie ligand homolngue nucleic acids. Patent Number: 6,426,216. Date of Patent: July 30, 2002.

Godowski P, Gurney A, Hillan KJ, Botstein D, **Goddard A**, Roy M, Ferrara N, Tumas D, Schwall R. NL4 Tie ligand homologue nucleic acid. Patent Number: 6,4137,770. Date of Patent: July 2, 2002.

Ashkenazi A, Fong S, **Goddard A**, Gurney AL, Napier MA, Tumas D, Wood WI. Nucleic acid encoding A-33 related antigen poly peptides. Patent Number: 6,410,708. Date of Patent:: Jun. 25, 2002.

Botstein DA, Cohen RL, **Goddard AD**, Gurney AL, Hillan KJ, Lawrence DA, Levine AJ, Pennica D, Roy MA and Wood WI. WISP polypeptides and nucleic acids encoding same. Patent Number: 6,387,657. Date nf Patent: May 14, 2002.

Goddard A, Godowski PJ and Gurney AL. Tie ligands. Patent Number: 6,372,491. Date of Patent: April 16, 2002.

Godowski PJ, Gurney AL, **Goddard A** and Hillan K. TIE ligand homolngue antibody. Patent Number: 6,350,450. Date of Patent: Feb. 26, 2002.

Fong S, Ferrara N, **Goddard** A, Godowski PJ, Gurney AL, Hillan K and Williams PM. Tie receptor tyrosine kinase ligand homologues. Patent Number: 6,348,351. Date of Patent: Feb. 10, 2002.

Goddard A, Godowski PJ and Gurney AL. Ligand homologues. Patent Number: 6,348,350. Date of Patent: Feb. 19, 2002.

Attie KM, Carlsson LMS, Gesundheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 6,207,640. Date of Patent: March 27, 2001.

Fong S, Ferrara N, **Goddard A**, Godowski PJ, Gurney AL, Hillan K and Williams PM. Nucleic acids encoding NL-3. Patent Number: 6,074,673. Date of Patent: June 13, 2000

Attie K, Carlsson LMS, Gesunheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,824,642. Date of Patent: October 20, 1998

Attie K, Carlsson LMS, Gesunheit N and **Goddard A**. Treatment of partial growth hormone insensitivity syndrome. Patent Number: 5,646,113. Date of Patent: July 8, 1997

Multiple additional provisional applications filed

PUBLICATIONS

Seshasayee D, Dowd P, Gu Q, Erickson S, **Goddard AD** Comparative sequence analysis of the *HER2* locus in mouse and man. Manuscript in preparation.

Abuzzahab MJ, **Goddard A**, Grigorescu F, Lautier C, Smith RJ and Chernausek SD. Human IGF-1 receptor mutations resulting in pre- and post-natal growth retardation. Manuscript in preparation.

Aggarwal S, Xie, M-H, Foster J, Frantz G, Stinson J, Corpuz RT, Simmons L, Hillan K, Yansura DG, Vandlen RL, **Goddard AD** and Gurney AL. FHFR, a novel receptor for the fibroblast growth factors. Manuscript submitted.

Adams SH, Chui C, Schilbach SL, Yu XX, **Goddard AD**, Grimaldi JC, Lee J, Dowd P, Colman S., Lewin DA. (2001) BFIT, a unique acyl-CoA thioesterase induced in thermogenic brown adipose tissue: Cloning, organization of the human gene, and assessment of a potential link to obesity. *Biochemical Journal* **360**: 135-142.

Lee J, Ho WH, Maruoka M, Corpuz RT, Baldwin DT, Foster JS, **Goddard AD**, Yansura DG, Vandlen RL, Wood WI, Gurney AL. (2001) IL-17E, a novel proinflammatory ligand for the IL-17 receptor homolog IL-17Rh1. *Journal of Biological Chemistry* **276**(2): 1660-1664.

Xie M-H, Aggarwal S, Ho W-H, Foster J, Zhang Z, Stinson J, Wood WI, **Goddard AD** and Gurney AL. (2000) Interleukin (IL)-22, a novel human cytokine that signals through the interferon-receptor related proteins CRF2-4 and IL-22R. *Journal of Biological Chemistry* **275**: 31335-31339.

Weiss GA, Watanabe CK, Zhong A, **Goddard A** and Sidhu SS. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. USA* **97**: 8050-8954.

Guo S, Yamaguchi Y, Schilbach S, Wada T.;Lee J, **Goddard A**, French D , Handa H, Rosenthal A. (2000) A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**: 366-369.

Yan M, Wang L-C, Hymowitz SG, Schilbach S, Lee J, **Goddard A**, de Vos AM, Gao WQ, Dixit VM. (2000) Two-amino acid molecular switch in an epithelial morphogen that regulates binding to two distinct receptors. *Science* **290**: 523-527.

Sehl PD, Tai JTN, Hillan KJ, Brown LA, **Goddard A**, Yang R, Jin H and Lowe DG. (2000) Application of cDNA microarrays in determining molecular phenotype in cardiac growth, development, and response to injury. *Circulation* **101**: 1990-1999.

Guo S, Brush J, Teraoka H, **Goddard A**, Wilson SW, Mullins MC and Rosenthal A. (1999) Development of noradrenergic neurons in the zebrafish hindbrain requires BMP, FGF8, and the homeodomain protein soulless/Phox2A. *Neuron* **24**: 555-566.

Stone D, Murone, M, Luoh, S, Ye W, Armanini P, Gurney A, Phillips HS, Brush, J, **Goddard A**, de Sauvage FJ and Rosenthal A. (1999) Characterization of the human suppressor of fused, a negative regulator of the zinc-finger transcription factor Gli. *J. Cell Sci.* **112**: 4437-4448.

Xie M-H, Holcomb I, Deuel B, Dowd P, Huang A, Vagts A, Foster J, Liang J, Brush J, Gu Q, Hillan K, **Goddard A** and Gurney, A.L. (1999) FGF-19, a novel fibroblast growth factor with unique specificity for FGFR4. *Cytokine* **11**: 729-735.

Yan M, Lee J, Schilbach S, **Goddard A** and Dixit V. (1999) mE10, a novel caspase recruitment domain-containing proapoptotic molecule. *J. Biol. Chem.* 274(15): 10287-10292.

Gurney AL, Marsters SA, Huang RM, Pitti RM, Mark DT, Baldwin DT, Gray AM, Dowd P, Brush J, Heldens S, Schow P, **Goddard AD**, Wood WI, Baker KP, Godowski PJ and Ashkenazi A. (1999) Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Current Biology* 9(4): 215-218.

Ridgway JBB, Ng E, Kern JA, Lee J, Brush J, **Goddard A** and Carter P. (1999) Identification of a human anti-CD55 single-chain Fv by subtractive panning of a phage library using tumor and nontumor cell lines. *Cancer Research* 59: 2718-2723.

Pitti RM, Marsters SA, Lawrence DA, Roy M, Kischkel FC, Dowd P, Huang A, Donahue CJ, Sherwood SW, Baldwin DT, Godowski PJ, Wood WI, Gurney AL, Hillan KJ, Cohen RL, **Goddard AD**, Botstein D and Ashkenazi A. (1998) Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* 396(6712): 699-703.

Pennica D, Swanson TA, Welsh JW, Roy MA, Lawrence DA, Lee J, Brush J, Taneyhill LA, Deuel B, Lew M, Watanabe C, Cohen RL, Melhem MF, Finley GG, Quirke P, **Goddard AD**, Hillan KJ, Gurney AL, Botstein D and Levine AJ. (1998) WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors. *Proc. Natl. Acad. Sci. USA.* 95(25): 14717-14722.

Yang RB, Mark MR, Gray A, Huang A, Xie MH, Zhang M, **Goddard A**, Wood WI, Gurney AL and Godowski PJ. (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature* 395(6699): 284-288.

Merchant AM, Zhu Z, Yuan JQ, **Goddard A**, Adams CW, Presta LG and Carter P. (1998) An efficient route to human bispecific IgG. *Nature Biotechnology* 16(7): 677-681.

Marsters SA, Sheridan JP, Pitti RM, Brush J, **Goddard A** and Ashkenazi A. (1998) Identification of a ligand for the death-domain-containing receptor Apo3. *Current Biology* 8(9): 525-528.

Xie J, Murone M, Luoh SM, Ryan A, Gu Q, Zhang C, Bonifas JM, Lam CW, Hynes M, **Goddard A**, Rosenthal A, Epstein EH Jr. and de Sauvage FJ. (1998) Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature.* 391(6662): 90-92.

Marsters SA, Sheridan JP, Pitti RM, Huang A, Skubatch M, Baldwin D, Yuan J, Gurney A, **Goddard AD**, Godowski P and Ashkenazi A. (1997) A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Current Biology.* 7(12): 1003-1006.

Hynes M, Stone DM, Dowd M, Pitts-Meek S, **Goddard A**, Gurney A and Rosenthal A. (1997) Control of cell pattern in the neural tube by the zinc finger transcription factor *Gli-1*. *Neuron* 19: 15–26.

Sheridan JP, Marsters SA, Pitti RM, Gurney A, Skubatch M, Baldwin D, Ramakrishnan L, Gray CL, Baker K, Wood WI, **Goddard AD**, Godowski P, and Ashkenazi A. (1997) Control of TRAIL-Induced Apoptosis by a Family of Signaling and Decoy Receptors. *Science* 277 (5327): 818-821.

Goddard AD, Dowd P, Chernausek S, Geffner M, Gertner J, Hintz R, Hopwood N, Kaplan S, Plotnick L, Rogol A, Rosenfield R, Saenger P, Mauras N, Hershkopf R, Angulo M and Attie, K. (1997) Partial growth hormone insensitivity: The role of growth hormone receptor mutations in idiopathic short stature. *J. Pediatr.* **131**: S51-55.

Klein RD, Sherman D, Ho WH, Stone D, Bennett GL, Moffat B, Vandlen R, Simmons L, Gu Q, Hongo JA, Devaux B, Poulsen K, Armanini M, Nozaki C, Asai N, **Goddard A**, Phillips H, Henderson CE, Takahashi M and Rosenthal A. (1997) A GPI-linked protein that interacts with Ret to form a candidate neurturin receptor. *Nature*. **387**(6634): 717-21.

Stone DM, Hynes M, Armanini M, Swanson TA, Gu Q, Johnson RL, Scott MP, Pennica D, **Goddard A**, Phillips H, Noll M, Hooper JE, de Sauvage F and Rosenthal A. (1996) The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**(6605): 129-34.

Marsters SA, Sheridan JP, Donahue CJ, Pitti RM, Gray CL, **Goddard AD**, Bauer KD and Ashkenazi A. (1996) Apo-3, a new member of the tumor necrosis factor receptor family, contains a death domain and activates apoptosis and NF-kappa B. *Current Biology* **6**(12): 1669-76.

Rothe M, Xiong J, Shu HB, Williamson K, **Goddard A** and Goeddel DV. (1996) I-TRAF is a novel TRAF-interacting protein that regulates TRAF-mediated signal transduction. *Proc. Natl. Acad. Sci. USA* **93**: 8241-8246.

Yang M, Luoh SM, **Goddard A**, Reilly D, Henzel W and Bass S. (1996) The bglX gene located at 47.8 min on the Escherichia coli chromosome encodes a periplasmic beta-glucosidase. *Microbiology* **142**: 1659-65.

Goddard AD and Black DM. (1996) Familial Cancer In Molecular Endocrinology of Cancer. Waxman, J. Ed. Cambridge University Press, Cambridge UK, pp.187-215.

Treanor JJS, Goodman L, de Sauvage F, Stone DM, Poulson KT, Beck CD, Gray C, Armanini MP, Pollocks RA, Hefti F, Phillips HS, **Goddard A**, Moore MW, Buj-Bello A, Davis AM, Asai N, Takahashi M, Vandlen R, Henderson CE and Rosenthal A. (1996) Characterization of a receptor for GDNF. *Nature* **382**: 80-83.

Klein RD, Gu Q, **Goddard A** and Rosenthal A. (1996) Selection for genes encoding secreted proteins and receptors. *Proc. Natl. Acad. Sci. USA* **93**: 7108-7113.

Winslow JW, Moran P, Valverde J, Shih A, Yuan JQ, Wong SC, Tsai SP, **Goddard A**, Henzel WJ, Hefti F and Caras I. (1995) Cloning of AL-1, a ligand for an Eph-related tyrosine kinase receptor involved in axon bundle formation. *Neuron* **14**: 973-981.

Bennett BD, Zeigler-FC, Gu Q, Fendly B, **Goddard AD**, Gillett N and Matthews W. (1995) Molecular cloning of a ligand for the EPH-related receptor protein-tyrosine kinase Htk. *Proc. Natl. Acad. Sci. USA* **92**: 1866-1870.

Huang X, Yuang J, **Goddard A**, Foulis A, James RF, Lernmark A, Pujol-Borrell R, Rabinovitch A, Somoza N and Stewart TA. (1995) Interferon expression in the pancreases of patients with type I diabetes. *Diabetes* **44**: 658-664.

Goddard AD, Yuan JQ, Fairbairn L, Dexter M, Borrow J, Kozak C and Solomon E. (1995) Cloning of the murine homolog of the leukemia-associated PML gene. *Mammalian Genome* **6**: 732-737.

Goddard AD, Covello R, Luoh SM, Clackson T, Attie KM, Gesundheit N, Rundle AC, Wells JA, Carlsson LMTI and The Growth Hormone Insensitivity Study Group. (1995) Mutations of the growth hormone receptor in children with idiopathic short stature. *N. Engl. J. Med.* **333**: 1093-1098.

Kuo SS, Moran P, Gripp J, Armanini M, Phillips HS, **Goddard A** and Caras IW. (1994) Identification and characterization of Batk, a predominently brain-specific non-receptor protein tyrosine kinase related to Csk. *J. Neurosci. Res.* **38**: 705-715.

Mark MR, Scadden DT, Wang Z, Gu Q, **Goddard A** and Godowski PJ. (1994) Rse, a novel receptor-type tyrosine kinase with homology to Axl/Ufo, is expressed at high levels in the brain. *Journal of Biological Chemistry* **269**: 10720-10728.

Borrow J, Shipley J, Howe K, Kiely F, **Goddard A**, Sheer D, Srivastava A, Antony AC, Fioretos T, Mitelman F and Solomon E. (1994) Molecular analysis of simple variant translocations in acute promyelocytic leukemia. *Genes Chromosomes Cancer* **9**: 234-243.

Goddard AD and Solomon E. (1993) Genetics of Cancer. *Adv. Hum. Genet.* **21**: 321-376.

Borrow J, **Goddard AD**, Gibbons B, Katz F, Swirsky D, Floretos T, Dube I, Winfield DA, Kingston J, Hagemeijer A, Rees JKH, Lister AT and Solomon E. (1992) Diagnosis of acute promyelocytic leukemia by RT-PCR: Detection of *PML-RARA* and *RARA-PML* fusion transcripts. *Br. J. Haematol.* **82**: 529-540.

Goddard AD, Borrow J and Solomon E. (1992) A previously uncharacterized gene, PML, is fused to the retinoic acid receptor alpha gene in acute promyelocytic leukemia. *Leukemia* **6 Suppl 3**: 117S–119S.

Zhu X, Dunn JM, **Goddard AD**, Squire JA, Becker A, Phillips RA and Gallie BL. (1992) Mechanisms of loss of heterozygosity in retinoblastoma. *Cytogenet. Cell. Genet.* **59**: 248-252.

Foulkes W, **Goddard A**, and Patel K. (1991) Retinoblastoma linked with Seascale [letter]. *British Med. J.* **302**: 409.

Goddard AD, Borrow J, Freemont PS and Solomon E. (1991) Characterization of a novel zinc finger gene disrupted by the t(15;17) in acute promyelocytic leukemia. *Science* **254**: 1371-1374.

Solomon E, Borrow J and **Goddard AD**. (1991) Chromosomal aberrations in cancer. *Science* **254**: 1153-1160.

Pajunen L, Jones TA, **Goddard A**, Sheer D, Solomon E, Pihlajaniemi T and Kivirikko KI. (1991) Regional assignment of the human gene coding for a multifunctional peptide (P4HB) acting as the β–subunit of prolyl-4-hydroxylase and the enzyme protein disulfide isomerase to 17q25. *Cytogenet. Cell. Genet.* **56**: 165-168.

Borrow J, Black DM, **Goddard AD**, Yagle MK, Frischauf A.-M and Solomon E. (1991) Construction and regional localization of a *Not*I linking library from human chromosome 17q. *Genomics* **10**: 477–480.

Borrow J, **Goddard AD**, Sheer D and Solomon E. (1990) Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17. *Science* **249**: 1577-1580.

Myers JC, Jones TA, Pohjolainen E-R, Kadri AS, **Goddard AD**, Sheer D, Solomon E and Pihlajaniemi T. (1990) Molecular cloning of 5(IV) collagen and assignment of the gene to the region of the region of the X-chromosome containing the Alport Syndrome locus. *Am. J. Hum. Genet.* 46: 1024-1033.

Gallie BL, Squire JA, **Goddard A**, Dunn JM, Canton M, Hinton D, Zhu X and Phillips RA. (1990) Mechanisms of oncogenesis in retinoblastoma. *Lab. Invest.* 62: 384-408.

Goddard AD, Phillips RA, Greger V, Passarge E, Hopping W, Gallie BL and Horsthemke B. (1990) Use of the RB1 cDNA as a diagnostic probe in retinoblastoma families. *Clinical Genetics* 37: 117-126.

Zhu XP, Dunn JM, Phillips RA, **Goddard AD**, Paton KE, Becker A and Gallie BL. (1989) Germline, but not somatic, mutations of the RB1 gene preferentially involve the paternal allele. *Nature* 340: 312-314.

Gallie BL, Dunn JM, **Goddard A**, Becker A and Phillips RA. (1988) Identification of mutations in the putative retinoblastoma gene. In Molecular Biology of The Eye: Genes, Vision and Ocular Disease. UCLA Symposia on Molecular and Cellular Biology, New Series, Volume 88. J. Piatigorsky, T. Shinohara and P.S. Zelenka, Eds. Alan R. Liss, Inc., New York, 1988, pp. 427-436.

Goddard AD, Balakier H, Canton M, Dunn J, Squire J, Reyes E, Becker A, Phillips RA and Gallie BL. (1988) Infrequent genomic rearrangement and normal expression of the putative RB1 gene in retinoblastoma tumors. *Mol. Cell. Biol.* 8: 2082-2089.

Squire J, Dunn J, **Goddard A**, Hoffman T, Musarella M, Willard HF, Becker AJ, Gallie BL and Phillips RA. (1988) Cloning of the esterase D gene: A polymorphic gene probe closely linked to the retinoblastoma locus on chromosome 13. *Proc. Natl. Acad. Sci.* USA 83: 6573-6577.

Squire J, **Goddard AD**, Canton M, Becker A, Phillips RA and Gallie BL (1986) Tumour induction by the retinoblastoma mutation is independent of N-*myc* expression. *Nature* 322: 555-557.

Goddard AD, Heddle JA, Gallie BL and Phillips RA. (1985) Radiation sensitivity of fibroblasts of bilateral retinoblastoma patients as determined by micronucleus induction *in vitro*. *Mutation Research* 152: 31-38.

HellerEhrman
ATTORNEYS

RESEARCH

SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi*, Gavin Dollinger[1], P. Scan Walsh and Robert Griffith

Roche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. [1]Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. *Corresponding author.

We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR[1] to clinical diagnostics are well known[2,3], it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization[5,6], gel electrophoresis with or without use of restriction digestion[7,8], HPLC[9], or capillary electrophoresis[10]. These methods are labor-intense, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of "carryover" false positives in subsequent testing[11].

These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.[12], developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.[13], developed an assay in which the endogenous 5' exonuclease assay of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to dsDNA[14-16]. As outlined in Figure 1, a prototypic PCR

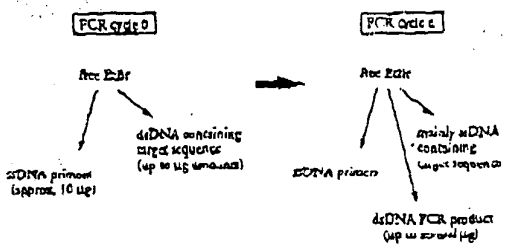


FIGURE 1 Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.

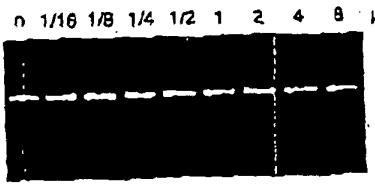


FIGURE 2 Gel electrophoresis of PCR amplification products of the human, nuclear gene, HLA DQα, made in the presence of increasing amounts of EtBr (up to 8 μg/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.

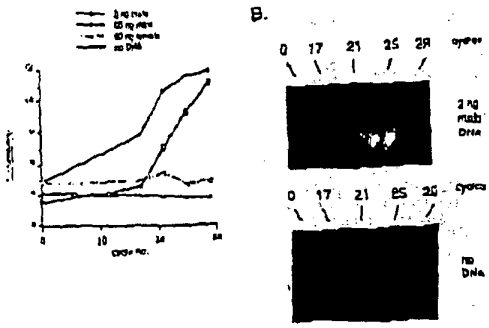


FIGURE 3 (A) Fluorescence measurements from PCRs that contain 0.5 μg/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photography of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ssDNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA[17] to micrograms per PCR[18]. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel before and after, or even continuously during, thermocycling.

RESULTS

PCR in the presence of EtBr. In order to assess the affect of EtBr in PCR, amplifications of the human HLA DQα gene[19] were performed with the dye present at concentrations from 0.06 to 8.0 μg/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5 μg/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

Detection of human Y-chromosome specific sequences. Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.5 μg/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome[20]. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

Detection of specific alleles of the human β-globin gene. In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5 μg/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human β-globin gene[21]. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3′ nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3′ nucleotide of the primer is complementary to the β-globin allele present[21,22].

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous wild-type β-globin individual (AA); from a heterozygous sickle β-globin individual (AS); and from a homozygous sickle β-globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a β-globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both β-globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for β-globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is *HIV*, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells[9]. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

FIGURE 4 UV photography of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human β-globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type allele, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCRs) for each input DNA.

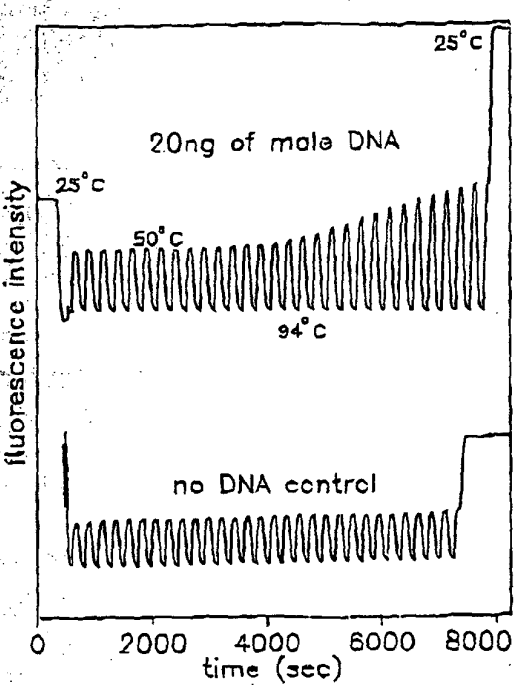


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.

DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube[8], and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins[8a]. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of 10^5 cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer[8a].

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format[8c]. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader[8c].

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiberoptics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

EXPERIMENTAL PROTOCOL

Human HLA-DQα gene amplifications containing EtBr. PCRs were set up in 100 μl volumes containing 10 mM Tris-HCl pH 8.3; 50 mM KCl; 4 mM $MgCl_2$; 2.5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQα gene specific oligonucleotide primers GH26 and GH27[19] and approximately 10^7 copies of DQα PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min. denaturation and 60°C for 30 sec. annealing and 72°C for 30 sec. extension.

Y-chromosome specific PCR. PCRs (100 μl total reaction volume) containing 0.5 μg/ml EtBr were prepared as described for HLA-DQα, except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2[20], and either 50 ng male, 60 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min. and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

Allele-specific human β-globin gene PCR. Amplifications of 100 μl volume using 0.5 μg/ml of EtBr were prepared as described for HLA-DQα above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/HBB4A (wild-type globin specific primers) or HGP5/HBB4S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.[21]. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the w.t. globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min. and 55°C for 1 min. using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.[21] to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 2#A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San Gabriel, CA).

Fluorescence measurement. Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 2 nm bandwidth with a GG 435 nm cut-off filter (Melles Griot, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a band width of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

Continuous fluorescence monitoring of PCR. Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute-epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 60°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration one

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.6 (SPEX) data system.

Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

References

1. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. CSHSQB 51:263–273.
2. White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. Trends Genet. 5:185–189.
3. Erlich, H. A., Gelfand, D. and Sninsky, J. J. 1991. Recent advances in the polymerase chain reaction. Science 252:1643–1651.
4. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491.
5. Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc. Nat. Acad. Sci. USA 86:6230–6234.
6. Kwok, S. Y., Mack, D. H., Mullis, K. B., Poiesz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using in vitro enzymatic amplification and oligomer cleavage detection. J. Virol. 61:1690–1694.
7. Chehab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemias. Nature 329:293–294.
8. Horn, G. T., Richards, B. and Klinger, K. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. Nuc. Acids Res. 10:2140.
9. Katz, E. D. and Ding, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. Biotechniques 8:548–553.
10. Heiger, D. N., Cohen, A. S. and Karger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. J. Chromatogr. 516:33–48.
11. Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. Nature 339:237–238.
12. Chehab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. Proc. Nat. Acad. Sci. USA 86:9178–9182.
13. Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5′ to 3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. Proc. Nat. Acad. Sci. USA 88:7276–7280.
14. Markovac, J., Roques, B. P. and Le Pecq, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. Anal. Biochem. 94:259–264.
15. Kapuscinski, J. and Szer, W. 1979. Interactions of 4′,6-diamidino-2-phenylindole with synthetic polynucleotides. Nuc. Acids Res. 6:3519–3534.
16. Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by ^{1}H NMR spectroscopy. Nuc. Acids Res. 18:3753–3762.
17. Li, H. H., Gyllensten, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335:414–417.
18. Abbott, M. A., Poiesz, B. J., Byrne, B. G., Kwok, S. Y., Sninsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified in vitro. J. Infect. Dis. 158:1158.
19. Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature 324:163–166.
20. Kogan, S. C., Doherty, M. and Gitschier, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. N. Engl. J. Med. 317:985–990.
21. Wu, D. Y., Ugozzoli, L., Pal, B. K. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. Proc. Nat. Acad. Sci. USA 86:2757–2760.
22. Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nuc. Acids Res. 18:999–1005.
23. Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. Submitted.
24. Higuchi, R. 1990. Using PCR to engineer DNA. p. 61–70. *In:* PCR Technology. H. A. Erlich (Ed.). Stockton Press. New York, N.Y.
25. Hall, L., Atwood, J. G., DiCesare, J., Katz, E., Pizzano, E., Williams, J. F. and Woudenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. Biotechniques 10:102–108, 110–112.
26. Tumosa, N. and Kahan, L. 1988. Fluorescent EIA screening of monoclonal antibodies to cell surface antigens. J. Immun. Meth. 116:59–68.

HellerEhrman
ATTORNEYS

Oligonucleotides with Fluorescent Dyes at Opposite Ends Provide a Quenched Probe System Useful for Detecting PCR Product and Nucleic Acid Hybridization

Kenneth J. Livak, Susan J.A. Flood, Jeffrey Marmaro, William Giusti, and Karin Deetz

Perkin-Elmer, Applied Biosystems Division, Foster City, California 94404

The 5′ nuclease PCR assay detects the accumulation of specific PCR product by hybridization and cleavage of a double-labeled fluorogenic probe during the amplification reaction. The probe is an oligonucleotide with both a reporter fluorescent dye and a quencher dye attached. An increase in reporter fluorescence intensity indicates that the probe has hybridized to the target PCR product and has been cleaved by the 5′→3′ nucleolytic activity of *Taq* DNA polymerase. In this study, probes with the quencher dye attached to an internal nucleotide were compared with probes with the quencher dye attached to the 3′-end nucleotide. In all cases, the reporter dye was attached to the 5′ end. All intact probes showed quenching of the reporter fluorescence. In general, probes with the quencher dye attached to the 3′-end nucleotide exhibited a larger signal in the 5′ nuclease PCR assay than the internally labeled probes. It is proposed that the larger signal is caused by increased likelihood of cleavage by *Taq* DNA polymerase when the probe is hybridized to a template strand during PCR. Probes with the quencher dye attached to the 3′-end nucleotide also exhibited an increase in reporter fluorescence intensity when hybridized to a complementary strand. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends can be used as homogeneous hybridiza-

A homogeneous assay for detecting the accumulation of specific PCR product that uses a double-labeled fluorogenic probe was described by Lee et al.[1] The assay exploits the 5′→3′ nucleolytic activity of *Taq* DNA polymerase[2,3] and is diagramed in Figure 1. The fluorogenic probe consists of an oligonucleotide with a reporter fluorescent dye, such as a fluorescein, attached to the 5′ end; and a quencher dye, such as a rhodamine, attached internally. When the fluorescein is excited by irradiation, its fluorescent emission will be quenched if the rhodamine is close enough to be excited through the process of fluorescence energy transfer (FET).[4-6] During PCR, if the probe is hybridized to a template strand, *Taq* DNA polymerase will cleave the probe because of its inherent 5′→3′ nucleolytic activity. If the cleavage occurs between the fluorescein and rhodamine dyes, it causes an increase in fluorescein fluorescence intensity because the fluorescein is no longer quenched. The increase in fluorescein fluorescence intensity indicates that the probe-specific PCR product has been generated. Thus, FET between a reporter dye and a quencher dye is critical to the performance of the probe in the 5′ nuclease PCR assay.

Quenching is completely dependent on the physical proximity of the two dyes.[6] Because of this, it has been assumed that the quencher dye must be attached near the 5′ end. Surprisingly, we have found that attaching a rhodamine dye at the 3′ end of a probe PCR assay. Furthermore, cleavage of this type of probe is not required to achieve some reduction in quenching. Oligonucleotides with a reporter dye on the 5′ end and a quencher dye on the 3′ end exhibit a much higher reporter fluorescence when double-stranded as compared with single-stranded. This should make it possible to use this type of double-labeled probe for homogeneous detection of nucleic acid hybridization.

MATERIALS AND METHODS

Oligonucleotides

Table 1 shows the nucleotide sequence of the oligonucleotides used in this study. Linker arm nucleotide (LAN) phosphoramidite was obtained from Glen Research. The standard DNA phosphoramidites, 6-carboxyfluorescein (6-FAM) phosphoramidite, 6-carboxytetramethylrhodamine succinimidyl ester (TAMRA NHS ester), and Phosphalink for attaching a 3′-blocking phosphate, were obtained from Perkin-Elmer, Applied Biosystems Division. Oligonucleotide synthesis was performed using an ABI model 394 DNA synthesizer (Applied Biosystems). Primer and complement oligonucleotides were purified using Oligo Purification Cartridges (Applied Biosystems). Double-labeled probes were synthesized with 6-FAM-labeled phosphoramidite at the 5′ end, LAN replacing one of the T's in the sequence, and Phosphalink at the 3′ end. Following deprotection and ethanol precipitation,

Research

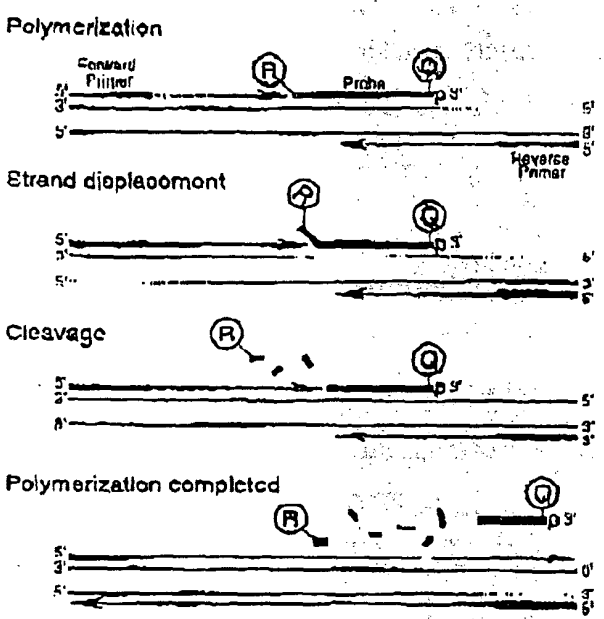


FIGURE 1 Diagram of 5′ nuclease assay. Stepwise representation of the 5′→3′ nucleolytic activity of *Taq* DNA polymerase acting on a fluorogenic probe during one extension phase of PCR.

mM Na-bicarbonate buffer (pH 9.0) at room temperature. Unreacted dye was removed by passage over a PD-10 Sephadex column. Finally, the double-labeled probe was purified by preparative high-performance liquid chromatography (HPLC) using an Aquapore C_{18} 220×4.6-mm column with 7-μm particle size. The column was developed with a 24-min linear gradient of 8–20% acetonitrile in 0.1 M TEAA (triethylamine acetate). Probes are named by designating the sequence from Table 1 and the position of the TAN–TAMRA moiety. For example, probe A1-7 has sequence A1 with TAN–TAMRA at nucleotide position 7 from the 5′ end.

PCR Systems

All PCR amplifications were performed in the Perkin-Elmer GeneAmp PCR System 9600 using 50-μl reactions that contained 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 400 μM dUTP, 0.5 unit of AmpErase uracil *N*-glycosylase (Perkin-Elmer), gene (nucleotides 2141–2435 in the sequence of Nakajima-Iijima et al.)[3] was amplified using primers AFP and ARP (Table 1), which are modified slightly from those of du Breuil et al.[4] Actin amplification reactions contained 4 mM $MgCl_2$, 20 ng of human genomic DNA, 50 nM A1 or A2 probe, and 300 nM each primer. The thermal regimen was 50°C (2 min), 95°C (10 min), 40 cycles of 95°C (20 sec), 60°C (1 min), and hold at 72°C. A 515-bp segment was amplified from a plasmid that consists of a segment of λ DNA (nucleotides 32,220–32,747) inserted in the *Sma*I site of vector pUC119. These reactions contained 5.5 mM $MgCl_2$, 1 ng of plasmid DNA, 50 nM P2 or P5 probe, 200 nM primer V119, and 300 nM primer R119. The thermal regimen was 50°C (2 min), 95°C (10 min), 25 cycles of 95°C (20 sec), 57°C (1 min), and hold at 72°C.

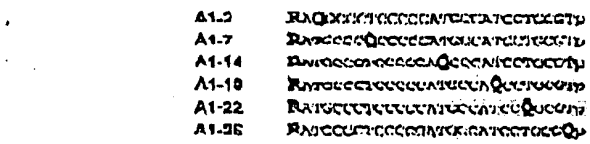
Fluorescence Detection

For each amplification reaction, a 40-μl aliquot of a sample was transferred to an individual well of a white, 96-well microtiter plate (Perkin-Elmer). Fluorescence was measured on the Perkin-Elmer Taq-Man LS-50B System, which consists of a luminescence spectrometer with plate reader assembly, a 488-nm excitation filter, and a 515-nm emission filter. Excitation was at 488 nm using a 5-nm slit width. Emission was measured at 518 nm for 6-FAM (the reporter or R value) and 582 nm for TAMRA (the quencher or Q value) using a 10-nm slit width. To determine the increase in reporter emission that is caused by cleavage of the probe during PCR, three normalizations are applied to the raw emission data. First, emission intensity of a buffer blank is subtracted for each wavelength. Second, emission intensity of the reporter is

TABLE 1 Sequences of Oligonucleotides

Name	Type	Sequence
V119	primer	ACCCACAGGAACTGATCACCACTC
R119	primer	ATGTCGCGTTCCGCCTCAGCGTTCTTC
P2	probe	TCCCATTACGATTCGTTCCGAACGACTp
P2C	complement	GTACTCGTTGGCAACGATCAGTAATGCCATG
P5	probe	CGCATTTGCTGGTATCTATTATAAAGCATp
P5C	complement	TTTATCGTTCTCATAGATAGCAGCAAATGCG
AFP	primer	TCACCCACACTGTGCCCATCTACGA
ARP	primer	CAGCTGGAAGGGCTCATTCCCAATGG
A1	probe	ATCCCCCTCCCCCATGCCATCCTCCCCTp
A1C	complement	AGAGGGGAGGATCGCATGGCCGAGGGGATAG
A2	probe	CGCCCTGGACTTCGAGCAAGAGATp
A2C	complement	CGATCTCTTGCTCGAAGTCCAGGGCGAG

For each oligonucleotide used in this study, the nucleic acid sequence is given, written in the 5′ → 3′ direction. There are three types of oligonucleotides: PCR primer, fluorogenic probe used



Probe	518 nm		582 nm		RQ⁻	RQ⁺	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A1-2	36.6 ± 2.1	32.7 ± 1.0	88.2 ± 6.0	88.0 ± 3.0	0.37 ± 0.01	0.33 ± 0.02	0.10 ± 0.02
A1-7	83.8 ± 4.3	305.1 ± 21.4	108.6 ± 6.0	110.3 ± 5.3	0.80 ± 0.02	0.58 ± 0.17	2.69 ± 0.18
A1-14	127.0 ± 4.9	433.5 ± 18.1	108.9 ± 5.3	93.1 ± 6.3	1.18 ± 0.03	4.34 ± 0.16	3.18 ± 0.15
A1-19	107.5 ± 17.9	400.7 ± 7.7	70.3 ± 7.4	78.9 ± 9.0	2.67 ± 0.05	5.08 ± 0.16	3.12 ± 0.16
A1-22	224.0 ± 9.4	480.9 ± 63.6	100.0 ± 4.0	96.2 ± 0.6	2.25 ± 0.03	5.02 ± 0.11	2.77 ± 0.12
A1-26	160.2 ± 0.9	424.1 ± 18.4	93.1 ± 5.4	90.7 ± 3.2	1.72 ± 0.02	5.01 ± 0.05	3.29 ± 0.05

FIGURE 2 Results of 5' nuclease assay comparing β-actin probes with TAMRA at different nucleotide positions. As described in Materials and Methods, PCR amplifications containing the indicated probes were performed, and the fluorescence emission was measured at 518 and 582 nm. Reported values are the average ± 1 s.d. for six reactions run without added template (no temp.) and six reactions run with template (+ temp.). The RQ ratio was calculated for each individual reaction and averaged to give the reported RQ⁻ and RQ⁺ values.

divided by the emission intensity of the quencher to give an RQ ratio for each reaction tube. This normalizes for well-to-well variations in probe concentration and fluorescence measurement. Finally, ΔRQ is calculated by subtracting the RQ value of the no-template control (RQ^-) from the RQ value for the complete reaction including template (RQ^+).

RESULTS

A series of probes with increasing distances between the fluorescein reporter and rhodamine quencher were tested to investigate the minimum and maximum spacing that would give an acceptable performance in the 5' nuclease PCR assay. These probes hybridize to a target sequence in the human β-actin gene. Figure 2 shows the results of an experiment in which these probes were included in PCR that amplified a segment of the β-actin gene containing the target sequence. Performance in the 5' nuclease PCR assay is monitored by the magnitude of ΔRQ, which is a measure of the increase in reporter fluorescence caused by PCR amplification of the probe target. Probe A1-2 has a ΔRQ value that is close to zero, indicating that the probe was not cleaved appreciably during the amplification reaction. This suggests that with the quencher dye on the second nucleotide from the 5' end, there is insufficient room for *Taq* polymerase to cleave efficiently between the reporter and quencher. The other five probes exhibited comparable ΔRQ values that are clearly different from zero. Thus, all five probes are being cleaved during PCR amplification resulting in a similar increase in reporter fluorescence. It should be noted that complete digestion of a probe produces a much larger increase in reporter fluorescence than that observed in Figure 2 (data not shown). Thus, even in reactions where amplification occurs, the majority of probe molecules remain uncleaved. It is mainly for this reason that the fluorescence intensity of the quencher dye TAMRA changes little with amplification of the target. This is what allows us to use the 582-nm fluorescence reading as a normalization factor.

The magnitude of RQ^- depends mainly on the quenching efficiency inherent in the specific structure of the probe and the purity of the oligonucleotide. Thus, the larger RQ^- values indicate that probes A1-14, A1-19, A1-22, and A1-26 probably have reduced quenching as compared with A1-7. Still, the degree of quenching is sufficient to detect a highly significant increase in reporter fluorescence when each of these probes is cleaved during PCR.

To further investigate the ability of TAMRA on the 3' end to quench 6-FAM on the 5' end, three additional pairs of probes were tested in the 5' nuclease PCR assay. For each pair, one probe has TAMRA attached to an internal nucleotide and the other has TAMRA attached to the 3' end nucleotide. The results are shown in Table 2. For all three sets, the probe with the 3' quencher exhibits a ΔRQ value that is considerably higher than for the probe with the internal quencher. The RQ^- values suggest that differences in quenching are not as great as those observed with some of the A1 probes. These results demonstrate that a quencher dye on the 3' end of an oligonucleotide can quench efficiently the

TABLE 2 Results of 5' Nuclease Assay Comparing Probes with TAMRA Attached to an Internal or 3'-terminal Nucleotide

Probe	518 nm		582 nm		RQ	RQ⁺	ΔRQ
	no temp.	+ temp.	no temp.	+ temp.			
A3-6	54.6 ± 3.2	84.8 ± 3.7	110.2 ± 6.4	113.6 ± 2.5	0.67 ± 0.02	0.73 ± 0.03	0.26 ± 0.04
A3-24	72.1 ± 2.9	236.5 ± 11.1	84.2 ± 4.0	90.2 ± 3.8	0.86 ± 0.02	2.62 ± 0.05	1.76 ± 0.05
P2-7	62.8 ± 4.4	384.0 ± 36.1	103.1 ± 6.9	120.4 ± 10.2	0.79 ± 0.02	3.19 ± 0.16	2.40 ± 0.16
P2-27	113.4 ± 6.6	556.4 ± 14.1	140.7 ± 8.5	118.7 ± 4.8	0.81 ± 0.01	4.68 ± 0.10	3.68 ± 0.10
P3-10	77.3 ± 6.5	244.4 ± 15.9	86.7 ± 4.3	95.8 ± 6.7	0.89 ± 0.06	2.55 ± 0.06	1.66 ± 0.08
P3-28	64.0 ± 3.2	333.6 ± 12.1	100.6 ± 8.1	94.7 ± 6.3	0.63 ± 0.02	3.54 ± 0.12	2.89 ± 0.13

* Amplifications were performed as described in Material and Methods and in the legend to Fig. 2.

fluorescence of a reporter dye on the 5′ end. The degree of quenching is sufficient for this type of oligonucleotide to be used as a probe in the 5′ nuclease PCR assay.

To test the hypothesis that quenching by a 3′ TAMRA depends on the flexibility of the oligonucleotide, fluorescence was measured for probes in the single-stranded and double-stranded states. Table 3 reports the fluorescence observed at 518 and 582 nm. The relative degree of quenching is assessed by calculating the RQ ratio. For probes with TAMRA 6–10 nucleotides from the 5′ end, there is little difference in the RQ values when comparing single-stranded with double-stranded oligonucleotides. The results for probes with TAMRA at the 3′ end are much different. For these probes, hybridization to a complementary strand causes a dramatic increase in RQ. We propose that this loss of quenching is caused by the rigid structure of double-stranded DNA, which prevents the 5′ and 3′ ends from being in proximity.

When TAMRA is placed toward the 3′ end, there is a marked Mg^{2+} effect on quenching. Figure 3 shows a plot of observed RQ values for the A1 series of probes as a function of Mg^{2+} concentration. With TAMRA attached near the 5′ end (probe A1-2 or A1-7), the RQ value at 0 mM Mg^{2+} is only slightly higher than RQ at 10 mM Mg^{2+}. For probes A1-19, A1-22, and A1-26, the RQ values at 0 mM Mg^{2+} are very high, indicating a much reduced quenching efficiency. For each of these probes, there is a marked decrease in RQ at 1 mM Mg^{2+} followed by a gradual decline as the Mg^{2+} concentration increases to 10 mM. Probe A1-14 shows an intermediate RQ value at 0 mM Mg^{2+} with a gradual decline at higher Mg^{2+} concentrations. In a low-salt environment with no Mg^{2+} present, a single-stranded oligonucleotide would be expected to adopt an extended conformation because of electrostatic repulsion. The binding of Mg^{2+} ions acts to shield the negative charge of the phosphate backbone so that the oligonucleotide can adopt conformations where the 3′ end is close to the 5′ end. Therefore, the observed Mg^{2+} effects support the notion that quenching of a 5′ reporter dye by TAMRA at or near the 3′ end depends on the flexibility of the oligonucleotide.

DISCUSSION

The striking finding of this study is that it seems the rhodamine dye TAMRA, placed at any position in an oligonucleotide, can quench the fluorescent emission of a fluorescein (6-FAM) placed at the 5′ end. This implies that a single-stranded, double-labeled oligonucleotide must be able to adopt conformations where the TAMRA is close to the 5′ end. It should be noted that the decay of 6-FAM in the excited state requires a certain amount of time. Therefore, what matters for quenching is not the average distance between 6-FAM and TAMRA but, rather, how close TAMRA can get to 6-FAM during the lifetime of the 6-FAM excited state. As long as the decay time of the excited state is relatively long compared with the molecular motions of the oligonucleotide, quenching can occur. Thus, we propose that TAMRA at the 3′ end, or any other position, can quench 6-FAM at the 5′ end because TAMRA is in proximity to 6-FAM often enough to be able to accept energy transfer from an excited 6-FAM.

Details of the fluorescence measurements remain puzzling. For example, Table 3 shows that hybridization of probes A1-26, A3-24, and P5-28 to their complementary strands not only causes a large increase in 6-FAM fluorescence at 518 nm but also causes a modest increase in TAMRA fluorescence at 582 nm. If TAMRA is being excited by energy transfer from quenched 6-FAM, then loss of quenching attributable to hybridization should cause a decrease in the fluorescence emission of TAMRA. The fact that the fluorescence emission of TAMRA increases indicates that the situation is more complex. For example, we have anecdotal evidence that the bases of the oligonucleotide, especially G, quench the fluorescence of both 6-FAM and TAMRA to some degree. When double-stranded, base-pairing may reduce the ability of the bases to quench. The primary factor causing the quenching of 6-FAM in an intact probe is the TAMRA dye. Evidence for the importance of TAMRA is that 6-FAM fluorescence remains relatively unchanged when probes labeled only with 6-FAM are used in the 5′ nuclease PCR assay (data not shown). Secondary effectors of fluorescence, both before and after cleavage of the probe, need to be explored further.

Regardless of the physical mechanism, the relative independence of position and quenching greatly simplifies the design of probes for the 5′ nuclease PCR assay. There are three main factors that determine the performance of a double-labeled fluorescent probe in the 5′ nuclease PCR assay. The first factor is the degree of quenching observed in the intact probe. This is characterized by the value of RQ′, which is the ratio of reporter to quencher fluorescent emis-

TABLE 3 Comparison of Fluorescence Emissions of Single-stranded and Double-stranded Fluorogenic Probes

	518 nm		582 nm		RQ	
Probe	ss	ds	ss	ds	ss	ds
A1-7	27.75	68.58	61.08	138.18	0.45	11.50
A1-26	43.41	809.38	53.50	93.86	0.81	5.43
A3-6	16.75	62.88	39.33	165.57	0.43	0.38
A3-24	30.05	578.64	67.77	140.25	0.45	3.71
P2-7	35.02	70.13	54.63	121.09	0.64	0.58
P2-27	20.80	220.47	65.10	61.13	0.61	5.25
P5-10	27.04	144.85	61.95	165.54	0.44	0.87
P5-28	33.66	462.20	92.30	104.41	0.46	4.43

(ss) Single-stranded. The fluorescence emissions at 518 or 582 nm for solutions containing a final concentration of 50 nM indicated probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and 10 mM $MgCl_2$. (ds) Double-stranded. The solutions contained, in addition, 100 nM A1C for probes A1-7 and A1-26, 100 nM A3C for probes A3-6 and A3-24, 100 nM P2C for probes P2-7 and P2-27, or 100 nM P5C for probes P5-10 and P5-28. Before the addition of $MgCl_2$, 150 µl of each sample was heated

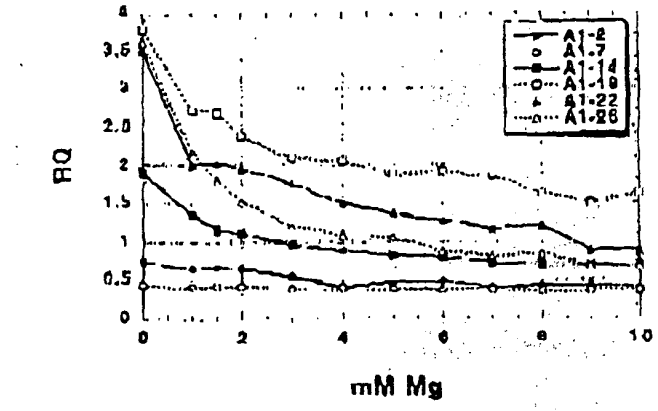


FIGURE 3 Effect of Mg^{++} concentration on RQ ratio for the A1 series of probes. The fluorescence emission intensity at 518 and 582 nm was measured for solutions containing 50 nM probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and varying amounts (0–10 mM) of $MgCl_2$. The calculated RQ ratios (518 nm intensity divided by 582 nm intensity) are plotted vs. $MgCl_2$ concentration (mM Mg). The key (*upper right*) shows the probes examined.

dyes used, spacing between reporter and quencher dyes, nucleotide sequence context effects, presence of structure or other factors that reduce flexibility of the oligonucleotide, and purity of the probe. The second factor is the efficiency of hybridization, which depends on probe T_m, presence of secondary structure in probe or template, annealing temperature, and other reaction conditions. The third factor is the efficiency at which *Taq* DNA polymerase cleaves the bound probe between the reporter and quencher dyes. This cleavage is dependent on sequence complementarity between probe and template as shown by the observation that mismatches in the segment between reporter and quencher dyes drastically reduce the cleavage of probe.[1]

The rise in RQ^- values for the A1 series of probes seems to indicate that the degree of quenching is reduced somewhat as the quencher is placed toward the 3' end. The lowest apparent quenching is observed for probe A1-19 (see Fig. 3) rather than for the probe where the TAMRA is at the 3' end (A1-26). This is understandable, as the conformation of the 3' end position would be expected to be less restricted than the conformation of an internal position. In effect, a quencher at the 3' end is freer to adopt conformations close to the 5' reporter dye than is an internally placed probe. The interpretation of RQ values is less clear-cut. The A3 probes show the same trend as A1, with the 3' TAMRA probe having a larger RQ^- than the internal TAMRA probe. For the P2 pair, both probes have about the same RQ^- value. For the P5 probes, the RQ for the 3' probe is less than for the internally labeled probe. Another factor that may explain some of the observed variation is that purity affects the RQ^- value. Although all probes are HPLC purified, a small amount of contamination with unquenched reporter can have a large effect on RQ.

Although there may be a modest effect on degree of quenching, the position of the quencher apparently can have a large effect on the efficiency of probe cleavage. The most drastic effect is observed with probe A1-2, where placement of the TAMRA on the second nucleotide reduces the efficiency of cleavage to almost zero. For the A3, P2, and P5 probes, ΔRQ is much greater for the 3' TAMRA probes as compared with the internal TAMRA probes. This is explained most easily by assuming that probes with TAMRA at the 3' end are more likely to be cleaved between reporter and quencher than are probes with TAMRA attached internally. For the A1 probes, the cleavage efficiency of probe A1-7 must already be quite high, as ΔRQ does not increase when the quencher is placed closer to the 3' end. This illustrates the importance of being able to use probes with a quencher on the 3' end in the 5' nuclease PCR assay. In this assay, an increase in the intensity of reporter fluorescence is observed only when the probe is cleaved between the reporter and quencher dyes. By placing the reporter and quencher dyes on the opposite ends of an oligonucleotide probe, any cleavage that occurs will be detected. When the quencher is attached to an internal nucleotide, sometimes the probe works well (A1-7) and other times not so well (A3-6). The relatively poor performance of probe A3-6 presumably means the probe is being cleaved 3' to the quencher rather than between the reporter and quencher. Therefore, the best chance of having a probe that reliably detects accumulation of PCR product in the 5' nuclease PCR assay is to use a probe with the reporter and quencher dyes on opposite ends.

Placing the quencher dye on the 3' end may also provide a slight benefit in terms of hybridization efficiency. The presence of a quencher attached to an internal nucleotide might be expected to disrupt base-pairing and reduce the T_m of a probe. In fact, a 2°C–3°C reduction in T_m has been observed for two probes with internally attached TAMRAs.[9] This disruptive effect would be minimized by placing the quencher at the 3' end. Thus, probes with 3' quenchers might exhibit slightly higher hybridization efficiencies than probes with internal quenchers.

The combination of increased cleavage and hybridization efficiencies means that probes with 3' quenchers probably will be more tolerant of mismatches between probe and target as compared with internally labeled probes. This tolerance of mismatches can be advantageous, as when trying to use a single probe to detect PCR-amplified products from samples of different species. Also, it means that cleavage of probe during PCR is less sensitive to alterations in annealing temperature or other reaction conditions. The one application where tolerance of mismatches may be a disadvantage is for allelic discrimination. Lee et al.[5] demonstrated that allele-specific probes were cleaved between reporter and quencher only when hybridized to a perfectly complementary target. This allowed them to distinguish the normal human cystic fibrosis allele from the *ΔF508* mutant. Their probes had TAMRA attached to the seventh nucleotide from

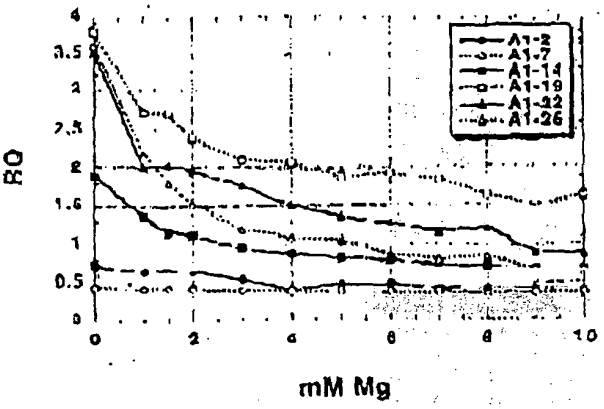


FIGURE 3 Effect of Mg^{2+} concentration on RQ ratio for the A1 series of probes. The fluorescence emission intensity at 518 and 582 nm was measured for solutions containing 50 nM probe, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and varying amounts (0–10 mM) of $MgCl_2$. The calculated RQ ratios (518 nm intensity divided by 582 nm intensity) are plotted vs. $MgCl_2$ concentration (mM Mg). The key (*upper right*) shows the probes examined.

dyes used, spacing between reporter and quencher dyes, nucleotide sequence context effects, presence of structure or other factors that reduce flexibility of the oligonucleotide, and purity of the probe. The second factor is the efficiency of hybridization, which depends on probe T_m, presence of secondary structure in probe or template, annealing temperature, and other reaction conditions. The third factor is the efficiency at which *Taq* DNA polymerase cleaves the bound probe between the reporter and quencher dyes. This cleavage is dependent on sequence complementarity between probe and template as shown by the observation that mismatches in the segment between reporter and quencher dyes drastically reduce the cleavage of probe.[1]

The rise in RQ values for the A1 series of probes seems to indicate that the degree of quenching is reduced somewhat as the quencher is placed toward the 3' end. The lowest apparent quenching is observed for probe A1-19 (see Fig. 3) rather than for the probe where the TAMRA is at the 3' end (A1-26). This is understandable, as the conformation of the 3' end position would be expected to be less restricted than the conformation of an internal position. In effect, a quencher at the 3' end is freer to adopt conformations close to the 5' reporter dye than is an internally placed

probes, the interpretation of RQ values is less clear-cut. The A3 probes show the same trend as A1, with the 3' TAMRA probe having a larger RQ^- than the internal TAMRA probe. For the P2 pair, both probes have about the same RQ value. For the P5 probes, the RQ^- for the 3' probe is less than for the internally labeled probe. Another factor that may explain some of the observed variation is that purity affects the RQ^- value. Although all probes are HPLC purified, a small amount of contamination with unquenched reporter can have a large effect on RQ.

Although there may be a modest effect on degree of quenching, the position of the quencher apparently can have a large effect on the efficiency of probe cleavage. The most drastic effect is observed with probe A1-2, where placement of the TAMRA on the second nucleotide reduces the efficiency of cleavage to almost zero. For the A3, P2, and P5 probes, ΔRQ is much greater for the 3' TAMRA probes as compared with the internal TAMRA probes. This is explained most easily by assuming that probes with TAMRA at the 3' end are more likely to be cleaved between reporter and quencher than are probes with TAMRA attached internally. For the A1 probes, the cleavage efficiency of probe A1-7 must already be quite high, as ΔRQ does not increase when the quencher is placed closer to the 3' end. This illustrates the importance of being able to use probes with a quencher on the 3' end in the 5' nuclease PCR assay. In this assay, an increase in the intensity of reporter fluorescence is observed only when the probe is cleaved between the reporter and quencher dyes. By placing the reporter and quencher dyes on the opposite ends of an oligonucleotide probe, any cleavage that occurs will be detected. When the quencher is attached to an internal nucleotide, sometimes the probe works well (A1-7) and other times not so well (A3-6). The relatively poor performance of probe A2-6 presumably means the probe is being cleaved 3' to the quencher rather than between the reporter and quencher. Therefore, the best chance of having a probe that reliably detects accumulation of PCR product in the 5' nuclease PCR assay is to use a probe with the reporter and quencher dyes on opposite ends.

Placing the quencher dye on the 3' end may also provide a slight benefit in terms of hybridization efficiency. The presence of a quencher attached to an internal nucleotide might be expected to disrupt base-pairing and reduce the T_m of a probe. In fact, a 2°C–3°C reduction in T_m has been observed for two probes with internally attached TAMRAs.[9] This disruptive effect would be minimized by placing the quencher at the 3' end. Thus, probes with 3' quenchers might exhibit slightly higher hybridization efficiencies than probes with internal quenchers.

The combination of increased cleavage and hybridization efficiencies means that probes with 3' quenchers probably will be more tolerant of mismatches between probe and target as compared with internally labeled probes. This tolerance of mismatches can be advantageous, as when trying to use a single probe to detect PCR-amplified products from samples of different species. Also, it means that cleavage of probe during PCR is less sensitive to alterations in annealing temperature or other reaction conditions. The one application where tolerance of mismatches may be a disadvantage is for allelic discrimination. Lee et al.[10] demonstrated that allele-specific probes were cleaved between reporter and quencher only when hybridized to a perfectly complementary target. This allowed them to distinguish the normal human cystic fibrosis allele from the *ΔF508* mutant. Their probes had TAMRA attached to the seventh nucleotide from

the 5' end and were designed so that any mismatches were between the reporter and quencher. Increasing the distance between reporter and quencher would lessen the disruptive effect of mismatches and allow cleavage of the probe on the incorrect target. Thus, probes with a quencher attached to an internal nucleotide may still be useful for allelic discrimination.

In this study loss of quenching upon hybridization was used to show that quenching by a 3' TAMRA is dependent on the flexibility of a single-stranded oligonucleotide. The increase in reporter fluorescence intensity, though, could also be used to determine whether hybridization has occurred or not. Thus, oligonucleotides with reporter and quencher dyes attached at opposite ends should also be useful as hybridization probes. The ability to detect hybridization in real time means that these probes could be used to measure hybridization kinetics. Also, this type of probe could be used to develop homogeneous hybridization assays for diagnostics or other applications. Bagwell et al.[10] describe just this type of homogeneous assay where hybridization of a probe causes an increase in fluorescence caused by a loss of quenching. However, they utilized a complex probe design that required adding nucleotides to both ends of the probe sequence to form two imperfect hairpins. The results presented here demonstrate that the simple addition of a reporter dye to one end of an oligonucleotide and a quencher dye to the other end generates a fluorogenic probe that can detect hybridization or PCR amplification.

ACKNOWLEDGMENTS

We acknowledge Lincoln McBride of Perkin-Elmer for his support and encouragement on this project and Mitch Winnik of the University of Toronto for helpful discussions on time-resolved fluorescence.

REFERENCES

1. Lee, L.G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* 21: 3761–3766.
2. [illegible] 1993. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* 260: 778–783.
3. Iyamichev, V., M.A.D. Brow, and J.E. Dahlberg. 1995. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* 260: 778–783.
4. Förster, V.Th. 1948. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys. (Leipzig)* 2: 55–75.
5. Lakowicz, J.R. 1983. Energy transfer. In *Principles of fluorescent spectroscopy*. pp. 303–339. Plenum Press, New York, NY.
6. Stryer, L. and R.P. Haugland. 1967. Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci.* 58: 719–726.
7. Nakajima-Iijima, S., H. Hamada, P. Reddy, and T. Kakunaga. 1985. Molecular structure of the human cytoplasmic beta-actin gene: Interspecies homology of sequences in the introns. *Proc. Natl. Acad. Sci.* 82: 6133–6137.
8. du Breuil, R.M., J.M. Patel, and R.V. Meredelow. 1993. Quantitation of β-actin-specific mRNA transcripts using semi-competitive PCR. *PCR Methods Applic.* 3: 57–59.
9. Livak, K.J. (unpubl.).
10. Bagwell, C.B., M.E. Munson, K.L. Christensen, and B.J. Lovett. 1994. A new homogeneous assay system for specific nucleic acid sequences. Poly-dA and poly-A detection. *Nucleic Acids Res.* 22: 2424–2425.

Received December 30, 1994; accepted in revised form March 6, 1995.

HellerEhrman
ATTORNEYS

RESEARCH

SIMULTANEOUS AMPLIFICATION AND DETECTION OF SPECIFIC DNA SEQUENCES

Russell Higuchi*, Gavin Dollinger[1], P. Sean Walsh and Robert Griffith

Roche Molecular Systems, Inc., 1400 53rd St., Emeryville, CA 94608. [1]Chiron Corporation, 1400 53rd St., Emeryville, CA 94608. *Corresponding author.

We have enhanced the polymerase chain reaction (PCR) such that specific DNA sequences can be detected without opening the reaction tube. This enhancement requires the addition of ethidium bromide (EtBr) to a PCR. Since the fluorescence of EtBr increases in the presence of double-stranded (ds) DNA an increase in fluorescence in such a PCR indicates a positive amplification, which can be easily monitored externally. In fact, amplification can be continuously monitored in order to follow its progress. The ability to simultaneously amplify specific DNA sequences and detect the product of the amplification both simplifies and improves PCR and may facilitate its automation and more widespread use in the clinic or in other situations requiring high sample throughput.

Although the potential benefits of PCR[1] to clinical diagnostics are well known[2,3], it is still not widely used in this setting, even though it is four years since thermostable DNA polymerases[4] made PCR practical. Some of the reasons for its slow acceptance are high cost, lack of automation of pre- and post-PCR processing steps, and false positive results from carryover-contamination. The first two points are related in that labor is the largest contributor to cost at the present stage of PCR development. Most current assays require some form of "downstream" processing once thermocycling is done in order to determine whether the target DNA sequence was present and has amplified. These include DNA hybridization[5,6], gel electrophoresis with or without use of restriction digestion[7,8], HPLC[9], or capillary electrophoresis[10]. These methods are labor-intense, have low throughput, and are difficult to automate. The third point is also closely related to downstream processing. The handling of the PCR product in these downstream processes increases the chances that amplified DNA will spread through the typing lab, resulting in a risk of "carryover" false positives in subsequent testing[11].

These downstream processing steps would be eliminated if specific amplification and detection of amplified DNA took place simultaneously within an unopened reaction vessel. Assays in which such different processes take place without the need to separate reaction components have been termed "homogeneous". No truly homogeneous PCR assay has been demonstrated to date, although progress towards this end has been reported. Chehab, et al.[12], developed a PCR product detection scheme using fluorescent primers that resulted in a fluorescent PCR product. Allele-specific primers, each with different fluorescent tags, were used to indicate the genotype of the DNA. However, the unincorporated primers must still be removed in a downstream process in order to visualize the result. Recently, Holland, et al.[13], developed an assay in which the endogenous 5' exonuclease assay of *Taq* DNA polymerase was exploited to cleave a labeled oligonucleotide probe. The probe would only cleave if PCR amplification had produced its complementary sequence. In order to detect the cleavage products, however, a subsequent process is again needed.

We have developed a truly homogeneous assay for PCR and PCR product detection based upon the greatly increased fluorescence that ethidium bromide and other DNA binding dyes exhibit when they are bound to dsDNA[14–16]. As outlined in Figure 1, a prototypic PCR

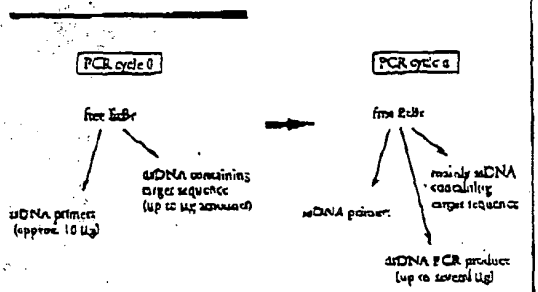


FIGURE 1 Principle of simultaneous amplification and detection of PCR product. The components of a PCR containing EtBr that are fluorescent are listed—EtBr itself, EtBr bound to either ssDNA or dsDNA. There is a large fluorescence enhancement when EtBr is bound to DNA and binding is greatly enhanced when DNA is double-stranded. After sufficient (n) cycles of PCR, the net increase in dsDNA results in additional EtBr binding, and a net increase in total fluorescence.

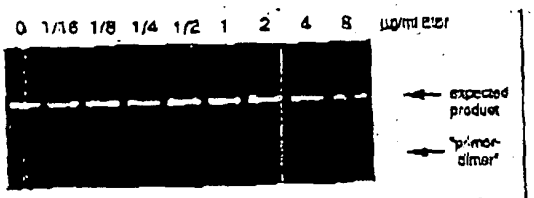


FIGURE 2 Gel electrophoresis of PCR amplification products of the human nuclear gene, HLA DQα, made in the presence of increasing amounts of EtBr (up to 8 μg/ml). The presence of EtBr has no obvious effect on the yield or specificity of amplification.

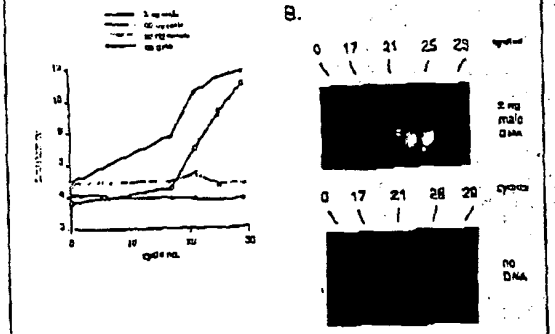


FIGURE 3 (A) Fluorescence measurements from PCRs that contain 0.6 μg/ml EtBr and that are specific for Y-chromosome repeat sequences. Five replicate PCRs were begun containing each of the DNAs specified. At each indicated cycle, one of the five replicate PCRs for each DNA was removed from thermocycling and its fluorescence measured. Units of fluorescence are arbitrary. (B) UV photography of PCR tubes (0.5 ml Eppendorf-style, polypropylene micro-centrifuge tubes) containing reactions, those starting from 2 ng male DNA and control reactions without any DNA, from (A).

begins with primers that are single-stranded DNA (ssDNA), dNTPs, and DNA polymerase. An amount of dsDNA containing the target sequence (target DNA) is also typically present. This amount can vary, depending on the application, from single-cell amounts of DNA[17] to micrograms per PCR[18]. If EtBr is present, the reagents that will fluoresce, in order of increasing fluorescence, are free EtBr itself, and EtBr bound to the single-stranded DNA primers and to the double-stranded target DNA (by its intercalation between the stacked bases of the DNA double-helix). After the first denaturation cycle, target DNA will be largely single-stranded. After a PCR is completed, the most significant change is the increase in the amount of dsDNA (the PCR product itself) of up to several micrograms. Formerly free EtBr is bound to the additional dsDNA, resulting in an increase in fluorescence. There is also some decrease in the amount of ssDNA primer, but because the binding of EtBr to ssDNA is much less than to dsDNA, the effect of this change on the total fluorescence of the sample is small. The fluorescence increase can be measured by directing excitation illumination through the walls of the amplification vessel before and after, or even continuously during, thermocycling.

RESULTS

PCR in the presence of EtBr. In order to assess the affect of EtBr in PCR, amplifications of the human HLA DQα gene[19] were performed with the dye present at concentrations from 0.06 to 8.0 μg/ml (a typical concentration of EtBr used in staining of nucleic acids following gel electrophoresis is 0.5 μg/ml). As shown in Figure 2, gel electrophoresis revealed little or no difference in the yield or quality of the amplification product whether EtBr was absent or present at any of these concentrations, indicating that EtBr does not inhibit PCR.

Detection of human Y-chromosome specific sequences. Sequence-specific, fluorescence enhancement of EtBr as a result of PCR was demonstrated in a series of amplifications containing 0.6 μg/ml EtBr and primers specific to repeat DNA sequences found on the human Y-chromosome[20]. These PCRs initially contained either 60 ng male, 60 ng female, 2 ng male human or no DNA. Five replicate PCRs were begun for each DNA. After 0, 17, 21, 24 and 29 cycles of thermocycling, a PCR for each DNA was removed from the thermocycler, and its fluorescence measured in a spectrofluorometer and plotted vs. amplification cycle number (Fig. 3A). The shape of this curve reflects the fact that by the time an increase in fluorescence can be detected, the increase in DNA is becoming linear and not exponential with cycle number. As shown, the fluorescence increased about three-fold over the background fluorescence for the PCRs containing human male DNA, but did not significantly increase for negative control PCRs, which contained either no DNA or human female DNA. The more male DNA present to begin with—60 ng versus 2 ng—the fewer cycles were needed to give a detectable increase in fluorescence. Gel electrophoresis on the products of these amplifications showed that DNA fragments of the expected size were made in the male DNA containing reactions and that little DNA synthesis took place in the control samples.

In addition, the increase in fluorescence was visualized by simply laying the completed, unopened PCRs on a UV transilluminator and photographing them through a red filter. This is shown in figure 3B for the reactions that began with 2 ng male DNA and those with no DNA.

Detection of specific alleles of the human β-globin gene. In order to demonstrate that this approach has adequate specificity to allow genetic screening, a detection of the sickle-cell anemia mutation was performed. Figure 4 shows the fluorescence from completed amplifications containing EtBr (0.5 μg/ml) as detected by photography of the reaction tubes on a UV transilluminator. These reactions were performed using primers specific for either the wild-type or sickle-cell mutation of the human β-globin gene[21]. The specificity for each allele is imparted by placing the sickle-mutation site at the terminal 3' nucleotide of one primer. By using an appropriate primer annealing temperature, primer extension—and thus amplification—can take place only if the 3' nucleotide of the primer is complementary to the β-globin allele present[21,22].

Each pair of amplifications shown in Figure 4 consists of a reaction with either the wild-type allele specific (left tube) or sickle-allele specific (right tube) primers. Three different DNAs were typed: DNA from a homozygous wild-type β-globin individual (AA); from a heterozygous sickle β-globin individual (AS); and from a homozygous sickle β-globin individual (SS). Each DNA (50 ng genomic DNA to start each PCR) was analyzed in triplicate (3 pairs

of reactions each). The DNA type was reflected in the relative fluorescence intensities in each pair of completed amplifications. There was a significant increase in fluorescence only where a β-globin allele DNA matched the primer set. When measured on a spectrofluorometer (data not shown), this fluorescence was about three times that present in a PCR where both β-globin alleles were mismatched to the primer set. Gel electrophoresis (not shown) established that this increase in fluorescence was due to the synthesis of nearly a microgram of a DNA fragment of the expected size for β-globin. There was little synthesis of dsDNA in reactions in which the allele-specific primer was mismatched to both alleles.

Continuous monitoring of a PCR. Using a fiber optic device, it is possible to direct excitation illumination from a spectrofluorometer to a PCR undergoing thermocycling and to return its fluorescence to the spectrofluorometer. The fluorescence readout of such an arrangement, directed at an EtBr-containing amplification of Y-chromosome specific sequences from 25 ng of human male DNA, is shown in Figure 5. The readout from a control PCR with no target DNA is also shown. Thirty cycles of PCR were monitored for each.

The fluorescence trace as a function of time clearly shows the effect of the thermocycling. Fluorescence intensity rises and falls inversely with temperature. The fluorescence intensity is minimum at the denaturation temperature (94°C) and maximum at the annealing/extension temperature (50°C). In the negative-control PCR, these fluorescence maxima and minima do not change significantly over the thirty thermocycles, indicating that there is little dsDNA synthesis without the appropriate target DNA, and there is little if any bleaching of EtBr during the continuous illumination of the sample.

In the PCR containing male DNA, the fluorescence maxima at the annealing/extension temperature begin to increase at about 4000 seconds of thermocycling, and continue to increase with time, indicating that dsDNA is being produced at a detectable level. Note that the fluorescence minima at the denaturation temperature do not significantly increase, presumably because at this temperature there is no dsDNA for EtBr to bind. Thus the course of the amplification is followed by tracking the fluorescence increase at the annealing temperature. Analysis of the products of these two amplifications by gel electrophoresis showed a DNA fragment of the expected size for the male DNA containing sample and no detectable DNA synthesis for the control sample.

DISCUSSION

Downstream processes such as hybridization to a sequence-specific probe can enhance the specificity of DNA detection by PCR. The elimination of these processes means that the specificity of this homogeneous assay depends solely on that of PCR. In the case of sickle-cell disease, we have shown that PCR alone has sufficient DNA sequence specificity to permit genetic screening. Using appropriate amplification conditions, there is little non-specific production of dsDNA in the absence of the appropriate target allele.

The specificity required to detect pathogens can be more or less than that required to do genetic screening, depending on the number of pathogens in the sample and the amount of other DNA that must be taken with the sample. A difficult target is HIV, which requires detection of a viral genome that can be at the level of a few copies per thousands of host cells[6]. Compared with genetic screening, which is performed on cells containing at least one copy of the target sequence, HIV detection requires both more specificity and the input of more total

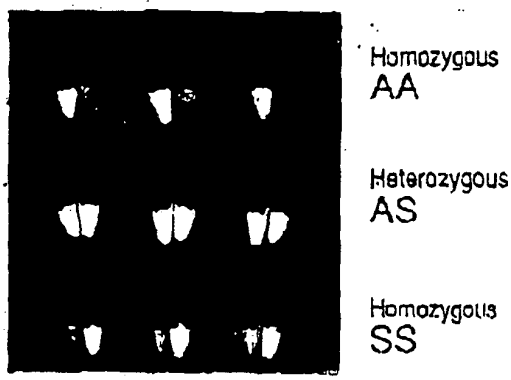


FIGURE 4 UV photography of PCR tubes containing amplifications using EtBr that are specific to wild-type (A) or sickle (S) alleles of the human β-globin gene. The left of each pair of tubes contains allele-specific primers to the wild-type allele, the right tube primers to the sickle allele. The photograph was taken after 30 cycles of PCR, and the input DNAs and the alleles they contain are indicated. Fifty ng of DNA was used to begin PCR. Typing was done in triplicate (3 pairs of PCRs) for each input DNA.

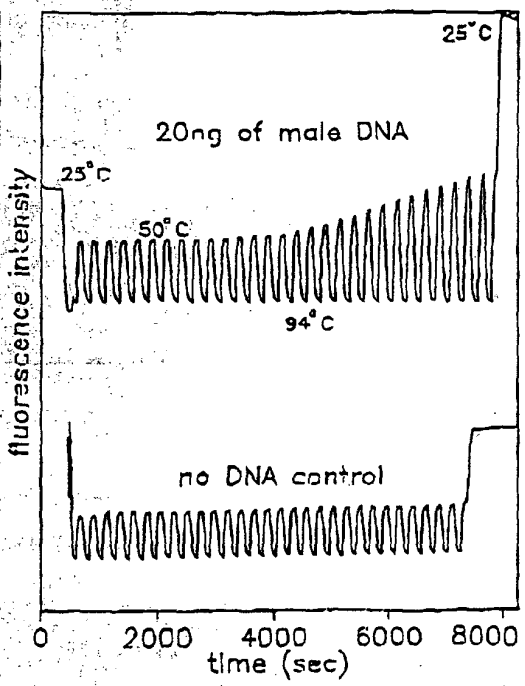


FIGURE 5 Continuous, real-time monitoring of a PCR. A fiber optic was used to carry excitation light to a PCR in progress and also emitted light back to a fluorometer (see Experimental Protocol). Amplification using human male-DNA specific primers in a PCR starting with 20 ng of human male DNA (top), or in a control PCR without DNA (bottom), were monitored. Thirty cycles of PCR were followed for each. The temperature cycled between 94°C (denaturation) and 50°C (annealing and extension). Note in the male DNA PCR, the cycle (time) dependent increase in fluorescence at the annealing/extension temperature.

DNA—up to microgram amounts—in order to have sufficient numbers of target sequences. This large amount of starting DNA in an amplification significantly increases the background fluorescence over which any additional fluorescence produced by PCR must be detected. An additional complication that occurs with targets in low copy-number is the formation of the "primer-dimer" artifact. This is the result of the extension of one primer using the other primer as a template. Although this occurs infrequently, once it occurs the extension product is a substrate for PCR amplification, and can compete with true PCR targets if those targets are rare. The primer-dimer product is of course dsDNA and thus is a potential source of false signal in this homogeneous assay.

To increase PCR specificity and reduce the effect of primer-dimer amplification, we are investigating a number of approaches, including the use of nested-primer amplifications that take place in a single tube[3], and the "hot-start", in which nonspecific amplification is reduced by raising the temperature of the reaction before DNA synthesis begins[22]. Preliminary results using these approaches suggest that primer-dimer is effectively reduced and it is possible to detect the increase in EtBr fluorescence in a PCR instigated by a single HIV genome in a background of 10^5 cells. With larger numbers of cells, the background fluorescence contributed by genomic DNA becomes problematic. To reduce this background, it may be possible to use sequence-specific DNA-binding dyes that can be made to preferentially bind PCR product over genomic DNA by incorporating the dye-binding DNA sequence into the PCR product through a 5' "add-on" to the oligonucleotide primer[23].

We have shown that the detection of fluorescence generated by an EtBr-containing PCR is straightforward, both once PCR is completed and continuously during thermocycling. The ease with which automation of specific DNA detection can be accomplished is the most promising aspect of this assay. The fluorescence analysis of completed PCRs is already possible with existing instrumentation in 96-well format[24]. In this format, the fluorescence in each PCR can be quantitated before, after, and even at selected points during thermocycling by moving the rack of PCRs to a 96-microwell plate fluorescence reader[25].

The instrumentation necessary to continuously monitor multiple PCRs simultaneously is also simple in principle. A direct extension of the apparatus used here is to have multiple fiberoptics transmit the excitation light and fluorescent emissions to and from multiple PCRs. The ability to monitor multiple PCRs continuously may allow quantitation of target DNA copy number. Figure 3 shows that the larger the amount of starting target DNA, the sooner during PCR a fluorescence increase is detected. Preliminary experiments (Higuchi and Dollinger, manuscript in preparation) with continuous monitoring have shown a sensitivity to two-fold differences in initial target DNA concentration.

Conversely, if the number of target molecules is known—as it can be in genetic screening—continuous monitoring may provide a means of detecting false positive and false negative results. With a known number of target molecules, a true positive would exhibit detectable fluorescence by a predictable number of cycles of PCR. Increases in fluorescence detected before or after that cycle would indicate potential artifacts. False negative results due to, for example, inhibition of DNA polymerase, may be detected by including within each PCR an inefficiently amplifying marker. This marker results in a fluorescence increase only after a large number of cycles—many more than are necessary to detect a true positive. If a sample fails to have a fluorescence increase after this many cycles, inhibition may be suspected. Since, in this assay, conclusions are drawn based on the presence or absence of fluorescence signal alone, such controls may be important. In any event, before any test based on this principle is ready for the clinic, an assessment of its false positive/false negative rates will need to be obtained using a large number of known samples.

In summary, the inclusion in PCR of dyes whose fluorescence is enhanced upon binding dsDNA makes it possible to detect specific DNA amplification from outside the PCR tube. In the future, instruments based upon this principle may facilitate the more widespread use of PCR in applications that demand the high throughput of samples.

EXPERIMENTAL PROTOCOL

Human HLA-DQα gene amplifications containing EtBr. PCRs were set up in 100 μl volumes containing 10 mM Tris-HCl, pH 8.3; 50 mM KCl; 4 mM $MgCl_2$; 2.5 units of *Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, CT); 20 pmole each of human HLA-DQα gene specific oligonucleotide primers GH26 and GH27[18] and approximately 10^5 copies of DQα PCR product diluted from a previous reaction. Ethidium bromide (EtBr; Sigma) was used at the concentrations indicated in Figure 2. Thermocycling proceeded for 20 cycles in a model 480 thermocycler (Perkin-Elmer Cetus, Norwalk, CT) using a "step-cycle" program of 94°C for 1 min. denaturation and 60°C for 40 sec. annealing and 72°C for 30 sec. extension.

Y-chromosome specific PCR. PCRs (100 μl total reaction volume) containing 0.5 μg/ml EtBr were prepared as described for HLA-DQα, except with different primers and target DNAs. These PCRs contained 15 pmole each male DNA-specific primers Y1.1 and Y1.2[19], and either 50 ng male, 50 ng female, 2 ng male, or no human DNA. Thermocycling was 94°C for 1 min. and 60°C for 1 min using a "step-cycle" program. The number of cycles for a sample were as indicated in Figure 3. Fluorescence measurement is described below.

Allele-specific, human β-globin gene PCR. Amplifications of 100 μl volume using 0.5 μg/ml of EtBr were prepared as described for HLA-DQα above except with different primers and target DNAs. These PCRs contained either primer pair HGP2/HB14A (wild-type globin specific primers) or HGP2/HB14S (sickle-globin specific primers) at 10 pmole each primer per PCR. These primers were developed by Wu et al.[21]. Three different target DNAs were used in separate amplifications—50 ng each of human DNA that was homozygous for the sickle trait (SS), DNA that was heterozygous for the sickle trait (AS), or DNA that was homozygous for the wild globin (AA). Thermocycling was for 30 cycles at 94°C for 1 min. and 55°C for 1 min. using a "step-cycle" program. An annealing temperature of 55°C had been shown by Wu et al.[21] to provide allele-specific amplification. Completed PCRs were photographed through a red filter (Wratten 88A) after placing the reaction tubes atop a model TM-36 transilluminator (UV-products San Gabriel, CA).

Fluorescence measurement. Fluorescence measurements were made on PCRs containing EtBr in a Fluorolog-2 fluorometer (SPEX, Edison, NJ). Excitation was at the 500 nm band with about 8 nm bandwidth with a GG 435 nm cut-off filter (Melles Griot, Inc., Irvine, CA) to exclude second-order light. Emitted light was detected at 570 nm with a bandwidth of about 7 nm. An OG 530 nm cut-off filter was used to remove the excitation light.

Continuous fluorescence monitoring of PCR. Continuous monitoring of a PCR in progress was accomplished using the spectrofluorometer and settings described above as well as a fiberoptic accessory (SPEX cat. no. 1950) to both send excitation light to, and receive emitted light from, a PCR placed in a well of a model 480 thermocycler (Perkin-Elmer Cetus). The probe end of the fiberoptic cable was attached with "5 minute-epoxy" to the open top of a PCR tube (a 0.5 ml polypropylene centrifuge tube with its cap removed) effectively sealing it. The exposed top of the PCR tube and the end of the fiberoptic cable were shielded from room light and the room lights were kept dimmed during each run. The monitored PCR was an amplification of Y-chromosome-specific repeat sequences as described above, except using an annealing/extension temperature of 50°C. The reaction was covered with mineral oil (2 drops) to prevent evaporation. Thermocycling and fluorescence measurement were started simultaneously. A time-base scan with a 10 second integration time

was used and the emission signal was ratioed to the excitation signal to control for changes in light-source intensity. Data were collected using the dm3000f, version 2.5 (SPEX) data system.

Acknowledgments

We thank Bob Jones for help with the spectrofluorometric measurements and Heatherbell Fong for editing this manuscript.

References

1. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. 1986. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. CSHSQB 51:263–273.
2. White, T. J., Arnheim, N. and Erlich, H. A. 1989. The polymerase chain reaction. Trends Genet. 5:185–189.
3. Erlich, H. A., Gelfand, D. and Sninsky, J. J. 1991. Recent advances in the polymerase chain reaction. Science 252:1643–1651.
4. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491.
5. Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc. Natl. Acad. Sci. USA 86:6230–6234.
6. Kwok, S. Y., Mack, D. H., Mullis, K. B., Poiesz, B. J., Ehrlich, G. D., Blair, D. and Friedman-Kien, A. S. 1987. Identification of human immunodeficiency virus sequences by using in vitro enzymatic amplification and oligomer cleavage detection. J. Virol. 61:1690–1694.
7. Chehab, F. F., Doherty, M., Cai, S. P., Kan, Y. W., Cooper, S. and Rubin, E. M. 1987. Detection of sickle cell anemia and thalassemias. Nature 329:293–294.
8. Horn, G. T., Richards, B. and Klinger, K. W. 1989. Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. Nuc. Acids Res. 17:2140.
9. Kain, K. D. and Dong, M. W. 1990. Rapid analysis and purification of polymerase chain reaction products by high-performance liquid chromatography. Biotechniques 8:546–555.
10. Huget, D. N., Cohen, A. S. and Karger, B. L. 1990. Separation of DNA restriction fragments by high performance capillary electrophoresis with low and zero crosslinked polyacrylamide using continuous and pulsed electric fields. J. Chromatogr. 516:33–48.
11. Kwok, S. Y. and Higuchi, R. G. 1989. Avoiding false positives with PCR. Nature 339:237–238.
12. Chehab, F. F. and Kan, Y. W. 1989. Detection of specific DNA sequences by fluorescence amplification: a color complementation assay. Proc. Natl. Acad. Sci. USA 86:9178–9182.
13. Holland, P. M., Abramson, R. D., Watson, R. and Gelfand, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. Proc. Natl. Acad. Sci. USA 88:7276–7280.
14. Markovits, J., Roques, B. P. and Le Pecq, J. B. 1979. Ethidium dimer: a new reagent for the fluorimetric determination of nucleic acids. Anal. Biochem. 94:259–264.
15. Kapuscinski, J. and Szer, W. 1979. Interactions of 3',6-diamidinephenylindole with synthetic polynucleotides. Nucl. Acids Res. 6:3519–3534.
16. Searle, M. S. and Embrey, K. J. 1990. Sequence-specific interaction of Hoechst 33258 with the minor groove of an adenine-tract DNA duplex studied in solution by ^{1}H NMR spectroscopy. Nuc. Acids Res. 18:3753–3762.
17. Li, H. H., Gyllensten, U. B., Cui, X. F., Saiki, R. K., Erlich, H. A. and Arnheim, N. 1988. Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335:414–417.
18. Abbott, M. A., Poiesz, B. J., Byrne, B. C., Kwok, S. Y., Sninsky, J. J. and Erlich, H. A. 1988. Enzymatic gene amplification: qualitative and quantitative methods for detecting proviral DNA amplified in vitro. J. Infect. Dis. 158:1158.
19. Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. 1986. Analysis of enzymatically amplified β-globin and HLA-DQα DNA with allele-specific oligonucleotide probes. Nature 324:163–166.
20. Kogan, S. C., Doherty, M. and Gitschier, J. 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. N. Engl. J. Med. 317:985–990.
21. Wu, D. Y., Ugozzoli, L., Pal, B. K. and Wallace, R. B. 1989. Allele-specific enzymatic amplification of β-globin genomic DNA for diagnosis of sickle cell anemia. Proc. Natl. Acad. Sci. USA 86:2757–2760.
22. Kwok, S., Kellogg, D. E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J. J. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. Nuc. Acids Res. 18:999–1005.
23. Chou, Q., Russell, M., Birch, D., Raymond, J. and Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. Submitted.
24. Higuchi, R. 1989. Using PCR to engineer DNA, p. 61–70. *In:* PCR Technology. H. A. Erlich (Ed.). Stockton Press, New York, N.Y.
25. Haff, L., Atwood, J. G., DiCesare, J., Katz, E., Picozza, E., Williams, J. F. and Woudenberg, T. 1991. A high-performance system for automation of the polymerase chain reaction. Biotechniques 10:102–103, 106–112.
26. Tumosa, N. and Kahan, L. 1986. Fluorescent EIA screening of monoclonal antibodies to cell surface antigens. J. Immun. Meth. 116:59–65.

HellerEhrman
ATTORNEYS

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 14717–14722, December 1998
Cell Biology, Medical Sciences

WISP genes are members of the connective tissue growth factor family that are up-regulated in Wnt-1-transformed cells and aberrantly expressed in human colon tumors

DIANE PENNICA*†, TODD A. SWANSON*, JAMES W. WELSH*, MARGARET A. ROY‡, DAVID A. LAWRENCE*, JAMES LEE‡, JENNIFER BRUSH‡, LISA A. TANEYHILL§, BETHANNE DEUEL‡, MICHAEL LEW¶, COLIN WATANABE‖, ROBERT L. COHEN*, MONA P. MELHEM**, GENE G. FINLEY**, PHIL QUIRKE††, AUDREY D. GODDARD‡, KENNETH J. HILLAN¶, AUSTIN L. GURNEY‡, DAVID BOTSTEIN§‡‡, AND ARNOLD J. LEVINE§

Departments of *Molecular Oncology, ‖Molecular Biology, Scientific Computing and **Pathology, Genentech Inc., 1 DNA Way, South San Francisco, CA 94080; **University of Pittsburgh School of Medicine, Veterans Administration Medical Center, Pittsburgh, PA 15240; ††University of Leeds, Leeds, LS2 9JT United Kingdom; ‡‡Department of Genetics, Stanford University, Palo Alto, CA 94305; and §Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by David Botstein and Arnold J. Levine, October 21, 1998

ABSTRACT Wnt family members are critical to many developmental processes, and components of the Wnt signaling pathway have been linked to tumorigenesis in familial and sporadic colon carcinomas. Here we report the identification of two genes, *WISP-1* and *WISP-2*, that are up-regulated in the mouse mammary epithelial cell line CS7MG transformed by Wnt-1, but not by Wnt-4. Together with a third related gene, *WISP-3*, these proteins define a subfamily of the connective tissue growth factor family. Two distinct systems demonstrated *WISP* induction to be associated with the expression of Wnt-1. These included (*i*) CS7MG cells infected with a Wnt-1 retroviral vector or expressing Wnt-1 under the control of a tetracycline repressible promoter, and (*ii*) Wnt-1 transgenic mice. The *WISP-1* gene was localized to human chromosome 8q24.1–8q24.3. *WISP-1* genomic DNA was amplified in colon cancer cell lines and in human colon tumors and its RNA overexpressed (2- to >30-fold) in 84% of the tumors examined compared with patient-matched normal mucosa. *WISP-3* mapped to chromosome 6q22–6q23 and also was overexpressed (4- to >40-fold) in 63% of the colon tumors analyzed. In contrast, *WISP-2* mapped to human chromosome 20q12–20q13 and its DNA was amplified, but RNA expression was reduced (2- to >30-fold) in 79% of the tumors. These results suggest that the *WISP* genes may be downstream of Wnt-1 signaling and that aberrant levels of *WISP* expression in colon cancer may play a role in colon tumorigenesis.

Wnt-1 is a member of an expanding family of cysteine-rich, glycosylated signaling proteins that mediate diverse developmental processes such as the control of cell proliferation, adhesion, cell polarity, and the establishment of cell fates (1, 2). Wnt-1 originally was identified as an oncogene activated by the insertion of mouse mammary tumor virus in virus-induced mammary adenocarcinomas (3, 4). Although Wnt-1 is not expressed in the normal mammary gland, expression of Wnt-1 in transgenic mice causes mammary tumors (5).

In mammalian cells, Wnt family members initiate signaling by binding to the seven-transmembrane spanning Frizzled receptors and recruiting the cytoplasmic protein Dishevelled (Dsh) to the cell membrane (1, 2, 6). Dsh then inhibits the kinase activity of the normally constitutively active glycogen synthase kinase-3β (GSK-3β) resulting in an increase in β-catenin levels. Stabilized β-catenin interacts with the transcription factor TCF/Lef1, forming a complex that appears in the nucleus and binds TCF/Lef1 target DNA elements to activate transcription (7, 8). Other experiments suggest that the adenomatous polyposis coli (APC) tumor suppressor gene also plays an important role in Wnt signaling by regulating β-catenin levels (9). APC is phosphorylated by GSK-3β, binds to β-catenin, and facilitates its degradation. Mutations in either APC or β-catenin have been associated with colon carcinomas and melanomas, suggesting these mutations contribute to the development of these types of cancer, implicating the Wnt pathway in tumorigenesis (1).

Although much has been learned about the Wnt signaling pathway over the past several years, only a few of the transcriptionally activated downstream components activated by Wnt have been characterized. Those that have been described cannot account for all of the diverse functions attributed to Wnt signaling. Among the candidate Wnt target genes are those encoding the nodal-related 3 gene, *Xnr3*, a member of the transforming growth factor (TGF)-β superfamily, and the homeobox genes, *engrailed*, *goosecoid*, *twin* (*Xtwn*), and *siamois* (2). A recent report also identifies c-*myc* as a target gene of the Wnt signaling pathway (10).

To identify additional downstream genes in the Wnt signaling pathway that are relevant to the transformed cell phenotype, we used a PCR-based cDNA subtraction strategy, suppression subtractive hybridization (SSH) (11), using RNA isolated from CS7MG mouse mammary epithelial cells and CS7MG cells stably transformed by a Wnt-1 retrovirus. Overexpression of Wnt-1 in this cell line is sufficient to induce a partially transformed phenotype, characterized by elongated and refractile cells that lose contact inhibition and form a multilayered array (12, 13). We reasoned that genes differentially expressed between these two cell lines might contribute to the transformed phenotype.

In this paper, we describe the cloning and characterization of two genes up-regulated in Wnt-1 transformed cells, *WISP-1* and *WISP-2*, and a third related gene, *WISP-3*. The *WISP* genes are members of the CCN family of growth factors, which includes connective tissue growth factor (CTGF), Cyr61, and *nov*, a family not previously linked to Wnt signaling.

MATERIALS AND METHODS

SSH. SSH was performed by using the PCR-Select cDNA Subtraction Kit (CLONTECH). Tester double-stranded

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: TGF, transforming growth factor; CTGF, connective tissue growth factor; SSH, suppression subtractive hybridization; VWC, von Willebrand factor type C module.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. AF100777, AF100778, AF100779, AF100780, and AF100781).

†To whom reprint requests should be addressed. e-mail: diane@gene.com.

cDNA was synthesized from 2 μg of poly(A)$^+$ RNA isolated from the C57MG/Wnt-1 cell line and driver cDNA from 2 μg of poly(A)$^+$ RNA from the parent C57MG cells. The subtracted cDNA library was subcloned into a pGEM-T vector for further analysis.

cDNA Library Screening. Clones encoding full-length mouse *WISP-1* were isolated by screening a λgt10 mouse embryo cDNA library (CLONTECH) with a 711-bp probe from the original partial clone S68 sequence corresponding to amino acids 128–169. Clones encoding full-length human *WISP-1* were isolated by screening λgt10 lung and fetal kidney cDNA libraries with the same probe at low stringency. Clones encoding full-length mouse and human *WISP-2* were isolated by screening a C57MG/Wnt-1 or human fetal lung cDNA library with a probe corresponding to nucleotides 1463–1512. Full-length cDNAs encoding *WISP-3* were cloned from human bone marrow and fetal kidney libraries.

Expression of Human *WISP* RNA. PCR amplification of first-strand cDNA was performed with human Multiple Tissue cDNA panels (CLONTECH) and 300 μM of each dNTP at 94°C for 1 sec, 62°C for 30 sec, 72°C for 1 min, for 22–32 cycles. *WISP* and glyceraldehyde-3-phosphate dehydrogenase primer sequences are available on request.

***In Situ* Hybridization.** ^{35}P-labeled sense and antisense riboprobes were transcribed from an 897-bp PCR product corresponding to nucleotides 601–1440 of mouse *WISP-1* or a 294-bp PCR product corresponding to nucleotides 82–375 of mouse *WISP-2*. All tissues were processed as described (40).

Radiation Hybrid Mapping. Genomic DNA from each hybrid in the Stanford G3 and Genebridge4 Radiation Hybrid Panels (Research Genetics, Huntsville, AL) and human and hamster control DNAs were PCR-amplified, and the results were submitted to the Stanford or Massachusetts Institute of Technology web servers.

Cell Lines, Tumors, and Mucosa Specimens. Tissue specimens were obtained from the Department of Pathology (University of Pittsburgh) for patients undergoing colon resection and from the University of Leeds, United Kingdom. Genomic DNA was isolated (Qiagen) from the pooled blood of 10 normal human donors, surgical specimens, and the following ATCC human cell lines: SW480, COLO 320DM, HT-29, WiDr, and SW403 (colon adenocarcinomas), SW620 (lymph node metastasis, colon adenocarcinoma), HCT 116 (colon carcinoma), SK-CO-1 (colon adenocarcinoma, ascites), and HM7 (a variant of ATCC colon adenocarcinoma cell line LS 174T). DNA concentration was determined by using Hoechst dye 33258 intercalation fluorimetry. Total RNA was prepared by homogenization in 7 M GuSCN followed by centrifugation over CsCl cushions or prepared by using RNAzol.

Gene Amplification and RNA Expression Analysis. Relative gene amplification and RNA expression of *WISP*s and *c-myc* in the cell lines, colorectal tumors, and normal mucosa were determined by quantitative PCR. Gene-specific primers and fluorogenic probes (sequences available on request) were designed and used to amplify and quantitate the genes. The relative gene copy number was derived by using the formula $2^{(\Delta Ct)}$ where ΔCt represents the difference in amplification cycles required to detect the *WISP* genes in peripheral blood lymphocyte DNA compared with colon tumor DNA or colon tumor RNA compared with normal mucosal RNA. The δ-method was used for calculation of the SE of the gene copy number or RNA expression level. The *WISP*-specific signal was normalized to that of the glyceraldehyde-3-phosphate dehydrogenase housekeeping gene. All TaqMan assay reagents were obtained from Perkin–Elmer Applied Biosystems.

RESULTS

Isolation of *WISP-1* and *WISP-2* by SSH. To identify Wnt-1-inducible genes, we used the technique of SSH using the mouse mammary epithelial cell line C57MG and C57MG cells that stably express Wnt-1 (11). Candidate differentially expressed cDNAs (1,384 total) were sequenced. Thirty-nine percent of the sequences matched known genes or homologues, 32% matched expressed sequence tags, and 29% had no match. To confirm that the transcript was differentially expressed, semiquantitative reverse transcription–PCR and Northern analysis were performed by using mRNA from the C57MG and C57MG/Wnt-1 cells.

Two of the cDNAs, *WISP-1* and *WISP-2*, were differentially expressed, being induced in the C57MG/Wnt-1 cell line, but not in the parent C57MG cells or C57MG cells overexpressing Wnt-4 (Fig. 1 *A* and *B*). Wnt-4, unlike Wnt-1, does not induce the morphological transformation of C57MG cells and has no effect on β-catenin levels (13, 14). Expression of *WISP-1* was up-regulated approximately 3-fold in the C57MG/Wnt-1 cell line and *WISP-2* by approximately 5-fold by both Northern analysis and reverse transcription–PCR.

An independent, but similar, system was used to examine *WISP* expression after Wnt-1 induction. C57MG cells expressing the *Wnt-1* gene under the control of a tetracycline-repressible promoter produce low amounts of Wnt-1 in the repressed state but show a strong induction of *Wnt-1* mRNA and protein within 24 hr after tetracycline removal (8). The levels of Wnt-1 and *WISP* RNA isolated from these cells at various times after tetracycline removal were assessed by quantitative PCR. Strong induction of Wnt-1 mRNA was seen as early as 10 hr after tetracycline removal. Induction of *WISP* mRNA (2- to 6-fold) was seen at 48 and 72 hr (data not shown). These data support our previous observations that show that *WISP* induction is correlated with Wnt-1 expression. Because the induction is slow, occurring after approximately 48 hr, the induction of *WISP*s may be an indirect response to Wnt-1 signaling.

cDNA clones of human *WISP-1* were isolated and the sequence compared with mouse *WISP-1*. The cDNA sequences of mouse and human *WISP-1* were 1,766 and 2,830 bp in length, respectively, and encode proteins of 367 aa, with predicted relative molecular masses of ≈40,000 (M_r 40 K). Both have hydrophobic N-terminal signal sequences, 38 conserved cysteine residues, and four potential N-linked glycosylation sites and are 84% identical (Fig. 2*A*).

Full-length cDNA clones of mouse and human *WISP-2* were 1,734 and 1,293 bp in length, respectively, and encode proteins of 251 and 250 aa, respectively, with predicted relative molecular masses of ≈27,000 (M_r 27 K) (Fig. 2*B*). Mouse and human *WISP-2* are 73% identical. Human *WISP-2* has no potential N-linked glycosylation sites, and mouse *WISP-2* has one at

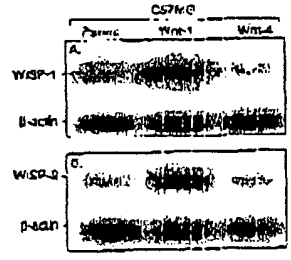


FIG. 1. *WISP-1* and *WISP-2* are induced by Wnt-1, but not Wnt-4, expression in C57MG cells. Northern analysis of *WISP-1* (*A*) and *WISP-2* (*B*) expression in C57MG, C57MG/Wnt-1, and C57MG/Wnt-4 cells. Poly(A)$^+$ RNA (2 μg) was subjected to Northern blot analysis and hybridized with a 70-bp mouse *WISP-1*-specific probe (amino acids 278–300) or a 190-bp *WISP-2*-specific probe (nucleotides 1438–1627) in the 3′ untranslated region. Blots were rehybridized with human β-actin probe.

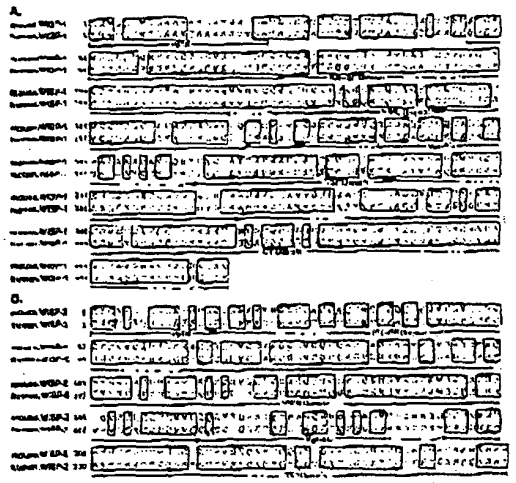


FIG. 2. Encoded amino acid sequence alignment of mouse and human *WISP-1* (*A*) and mouse and human *WISP-2* (*B*). The potential signal sequence, insulin-like growth factor-binding protein (IGF-BP), VWC, thrombospondin (TSP), and C-terminal (CT) domains are underlined.

position 197. *WISP-2* has 28 cysteine residues that are conserved among the 38 cysteines found in *WISP-1*.

Identification of *WISP-3*. To search for related proteins, we screened expressed sequence tag (EST) databases with the WISP-1 protein sequence and identified several ESTs as potentially related sequences. We identified a homologous protein that we have called WISP-3. A full-length human *WISP-3* cDNA of 1,371 bp was isolated corresponding to those ESTs that encode a 354-aa protein with a predicted molecular mass of 39,293. WISP-3 has two potential N-linked glycosylation sites and 36 cysteine residues. An alignment of the three human WISP proteins shows that WISP-1 and WISP-3 are the most similar (42% identity), whereas WISP-2 has 37% identity with WISP-1 and 32% identity with WISP-3 (Fig. 3*A*).

***WISPs* Are Homologous to the CTGF Family of Proteins.** Human *WISP-1*, *WISP-2*, and *WISP-3* are novel sequences; however, mouse *WISP-1* is the same as the recently identified *Elm1* gene. *Elm1* is expressed in low, but not high, metastatic mouse melanoma cells, and suppresses the *in vivo* growth and metastatic potential of K-1735 mouse melanoma cells (15). Human and mouse *WISP-2* are homologous to the recently described rat gene, *rCop-1* (16). Significant homology (36–44%) was seen to the CCN family of growth factors. This family includes three members, CTGF, Cyr61, and the protooncogene *nov*. CTGF is a chemotactic and mitogenic factor for fibroblasts that is implicated in wound healing and fibrotic disorders and is induced by TGF-β (17). Cyr61 is an extracellular matrix signaling molecule that promotes cell adhesion, proliferation, migration, angiogenesis, and tumor growth (18, 19). *nov* (nephroblastoma overexpressed) is an immediate early gene associated with quiescence and found altered in Wilms tumors (20). The proteins of the CCN family share functional, but not sequence, similarity to Wnt-1. All are secreted, cysteine-rich heparin binding glycoproteins that associate with the cell surface and extracellular matrix.

WISP proteins exhibit the modular architecture of the CCN family, characterized by four conserved cysteine-rich domains (Fig. 3*B*) (21). The N-terminal domain, which includes the first 12 cysteine residues, contains a consensus sequence (GCGC-CXXC) conserved in most insulin-like growth factor (IGF)-

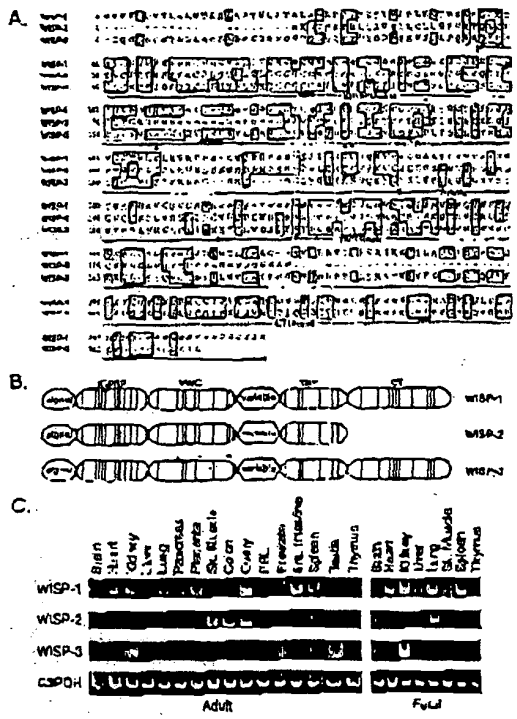


FIG. 3. (*A*) Encoded amino acid sequence alignment of human WISPs. The cysteine residues of WISP-1 and WISP-2 that are not present in WISP-3 are indicated with a dot. (*B*) Schematic representation of the WISP proteins showing the domain structure and cysteine residues (vertical lines). The four cysteine residues in the VWC domain that are absent in WISP-3 are indicated with a dot. (*C*) Expression of *WISP* mRNA in human tissues. PCR was performed on human multiple-tissue cDNA panels (CLONTECH) from the indicated adult and fetal tissues.

binding proteins (BP). This sequence is conserved in WISP-2 and WISP-3, whereas WISP-1 has a glutamine in the third position instead of a glycine. CTGF recently has been shown to specifically bind IGF (22) and a truncated *nov* protein lacking the IGF-BP domain is oncogenic (23). The von Willebrand factor type C module (VWC), also found in certain collagens and mucins, covers the next 10 cysteine residues, and is thought to participate in protein complex formation and oligomerization (24). The VWC domain of WISP-3 differs from all CCN family members described previously, in that it contains only six of the 10 cysteine residues (Fig. 3 *A* and *B*). A short variable region follows the VWC domain. The third module, the thrombospondin (TSP) domain is involved in binding to sulfated glycoconjugates and contains six cysteine residues and a conserved WSxCSxxCG motif first identified in thrombospondin (25). The C-terminal (CT) module containing the remaining 10 cysteines is thought to be involved in dimerization and receptor binding (26). The CT domain is present in all CCN family members described to date but is absent in WISP-2 (Fig. 3 *A* and *B*). The existence of a putative signal sequence and the absence of a transmembrane domain suggest that WISPs are secreted proteins, an observation supported by an analysis of their expression and secretion from mammalian cell and baculovirus cultures (data not shown).

Expression of *WISP* mRNA in Human Tissues. Tissue-specific expression of human *WISPs* was characterized by PCR

analysis on adult and fetal multiple tissue cDNA panels. *WISP-1* expression was seen in the adult heart, kidney, lung, pancreas, placenta, ovary, small intestine, and spleen (Fig. 3*C*). Little or no expression was detected in the brain, liver, skeletal muscle, colon, peripheral blood leukocytes, prostate, testis, or thymus. *WISP-2* had a more restricted tissue expression and was detected in adult skeletal muscle, colon, ovary, and fetal lung. Predominant expression of *WISP-3* was seen in adult kidney and testis and fetal kidney. Lower levels of *WISP-3* expression were detected in placenta, ovary, prostate, and small intestine.

***In Situ* Localization of *WISP-1* and *WISP-2*.** Expression of *WISP-1* and *WISP-2* was assessed by *in situ* hybridization in mammary tumors from Wnt-1 transgenic mice. Strong expression of *WISP-1* was observed in stromal fibroblasts lying within the fibrovascular tumor stroma (Fig. 4 *A–D*). However, low-level *WISP-1* expression also was observed focally within tumor cells (data not shown). No expression was observed in normal breast. Like *WISP-1*, *WISP-2* expression also was seen in the tumor stroma in breast tumors from Wnt-1 transgenic animals (Fig. 4 *E–H*). However, *WISP-2* expression in the stroma was in spindle-shaped cells adjacent to capillary vessels, whereas

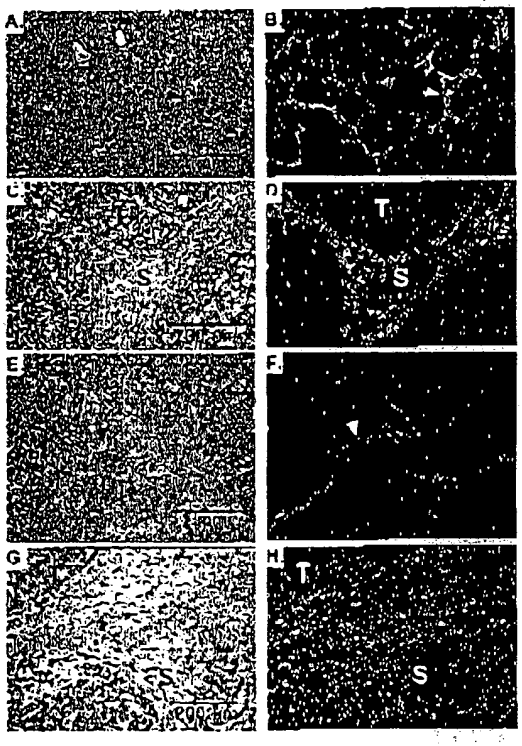


FIG. 4. (*A*, *C*, *E*, and *G*) Representative hematoxylin/eosin-stained images from breast tumors in Wnt-1 transgenic mice. The corresponding dark-field images showing *WISP-1* expression are shown in *B* and *D*. The tumor is a moderately well-differentiated adenocarcinoma showing evidence of adenoid cystic change. At low power (*A* and *B*), expression of *WISP-1* is seen in the delicate branching fibrovascular tumor stroma (arrowhead). At higher magnification, expression is seen in the stromal(s) fibroblasts (*C* and *D*), and tumor cells are negative. Focal expression of *WISP-1*, however, was observed in tumor cells in some areas. Images of *WISP-3* expression are shown in *E–H*. At low power (*E* and *F*), expression of WISP-2 is seen in cells lying within the fibrovascular tumor stroma. At higher magnification, these cells appeared to be adjacent to capillary vessels whereas tumor cells are negative (*G* and *H*).

the predominant cell type expressing *WISP-1* was the stromal fibroblasts.

Chromosome Localization of the *WISP* Genes. The chromosomal location of the human *WISP* genes was determined by radiation hybrid mapping panels. *WISP-1* is approximately 3.48 cR from the meiotic marker AFM259xc5 [logarithm of odds (lod) score 16.3] on chromosome 8q24.1 to 8q24.3, in the same region as the human locus of the *nov*H family member (27) and roughly 4 Mbs distal to *c-myc* (28). Preliminary fine mapping indicates that *WISP-1* is located near D8S1712 STS. *WISP-2* is linked to the marker SHGC-33922 (lod = 1,000) on chromosome 20q12–20q13.1. Human *WISP-3* mapped to chromosome 6q22–6q23 and is linked to the marker AFM211ze5 (lod = 1,000). *WISP-3* is approximately 18 Mbs proximal to CTGF and 23 Mbs proximal to the human cellular oncogene *MYB* (27, 29).

Amplification and Aberrant Expression of *WISPs* in Human Colon Tumors. Amplification of protooncogenes is seen in many human tumors and has etiological and prognostic significance. For example, in a variety of tumor types, *c-myc* amplification has been associated with malignant progression and poor prognosis (30). Because *WISP-1* resides in the same general chromosomal location (8q24) as *c-myc*, we asked whether it was a target of gene amplification, and, if so, whether this amplification was independent of the *c-myc* locus. Genomic DNA from human colon cancer cell lines was assessed by quantitative PCR and Southern blot analysis. (Fig. 5 *A* and *B*). Both methods detected similar degrees of *WISP-1* amplification. Most cell lines showed significant (2- to 4-fold) amplification, with the HT-29 and WiDr cell lines demonstrating an 8-fold increase. Significantly, the pattern of amplification observed did not correlate with that observed for *c-myc*, indicating that the *c-myc* gene is not part of the amplicon that involves the *WISP-1* locus.

We next examined whether the *WISP* genes were amplified in a panel of 25 primary human colon adenocarcinomas. The relative *WISP* gene copy number in each colon tumor DNA was compared with pooled normal DNA from 10 donors by quantitative PCR (Fig. 6). The copy number of *WISP-1* and *WISP-2* was significantly greater than one, approximately 2-fold for *WISP-1* in about 60% of the tumors and 3- to 4-fold for *WISP-2* in 92% of the tumors ($P < 0.001$ for each). The copy number for *WISP-3* was indistinguishable from one ($P = 0.166$). In addition, the copy number of *WISP-2* was significantly higher than that of *WISP-1* ($P < 0.001$).

The levels of *WISP* transcripts in RNA isolated from 19 adenocarcinomas and their matched normal mucosa were

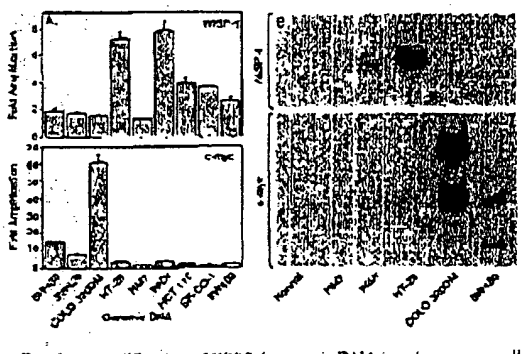


FIG. 5. Amplification of *WISP-1* genomic DNA in colon cancer cell lines (*A*) Amplification in cell line DNA was determined by quantitative PCR. (*B*) Southern blots containing genomic DNA (10 μg) digested with *Eco*RI (*WISP-1*) or *Xba*I (*c-myc*) were hybridized with a 100-bp human *WISP-1* probe (amino acids 186–219) or a human *c-myc* probe (located at bp 1901–2000). The *WISP* and *myc* genes are detected in normal human genomic DNA after a longer film exposure.

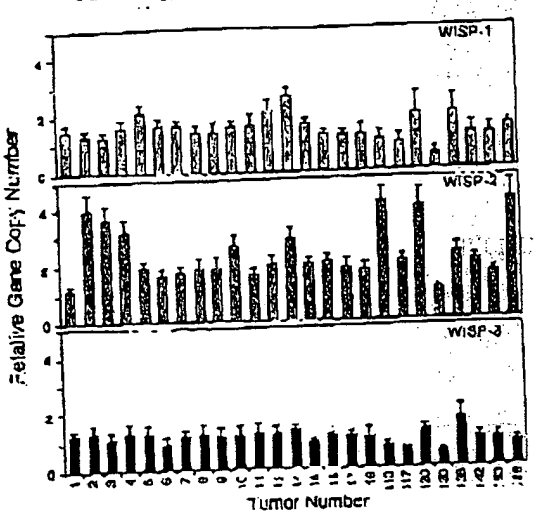


FIG. 6. Genomic amplification of *WISP* genes in human colon tumors. The relative gene copy number of the *WISP* genes in 25 adenocarcinomas was assayed by quantitative PCR, by comparing DNA from primary human tumors with pooled DNA from 10 healthy donors. The data are means ± SEM from one experiment done in triplicate. The experiment was repeated at least three times.

assessed by quantitative PCR (Fig. 7). The level of *WISP-1* RNA present in tumor tissue varied but was significantly increased (2- to >25-fold) in 84% (16/19) of the human colon tumors examined compared with normal adjacent mucosa. Four of 19 tumors showed greater than 10-fold overexpression. In contrast, in 79% (15/19) of the tumors examined, *WISP-2* RNA expression was significantly lower in the tumor than the mucosa. Similar to *WISP-1*, *WISP-3* RNA was overexpressed in 63% (12/19) of the colon tumors compared with the normal

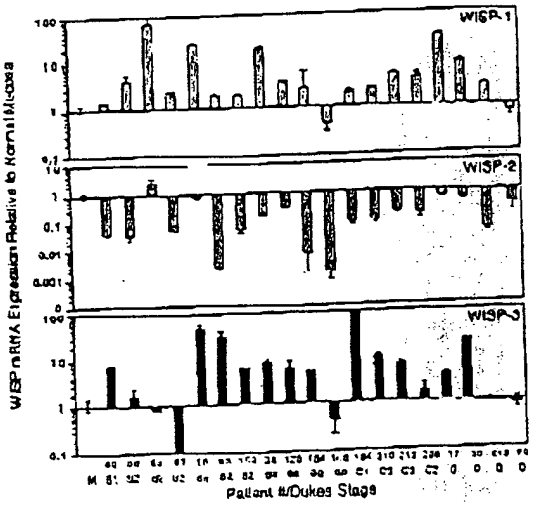


FIG. 7. *WISP* RNA expression in primary human colon tumors relative to expression in normal mucosa from the same patient. Expression of *WISP* mRNA in 19 adenocarcinomas was assayed by quantitative PCR. The Dukes stage of the tumor is listed under the sample number. The data are means ± SEM from one experiment done in triplicate. The experiment was repeated at least twice.

mucosa. The amount of overexpression of *WISP-3* ranged from 4- to >40-fold.

DISCUSSION

One approach to understanding the molecular basis of cancer is to identify differences in gene expression between cancer cells and normal cells. Strategies based on assumptions that steady-state mRNA levels will differ between normal and malignant cells have been used to clone differentially expressed genes (31). We have used a PCR-based selection strategy, SSH, to identify genes selectively expressed in CS7MG mouse mammary epithelial cells transformed by Wnt-1.

Three of the genes isolated, *WISP-1*, *WISP-2*, and *WISP-3*, are members of the CCN family of growth factors, which includes CTGF, Cyr61, and *nov*, a family not previously linked to Wnt signaling.

Two independent experimental systems demonstrated that *WISP* induction was associated with the expression of Wnt-1. The first was CS7MG cells infected with a Wnt-1 retroviral vector or CS7MG cells expressing Wnt-1 under the control of a tetracycline-repressible promoter, and the second was in Wnt-1 transgenic mice, where breast tissue expresses Wnt-1, whereas normal breast tissue does not. No *WISP* RNA expression was detected in mammary tumors induced by polyoma virus middle T antigen (data not shown). These data suggest a link between Wnt-1 and *WISPs* in that in these two situations, *WISP* induction was correlated with Wnt-1 expression.

It is not clear whether the *WISPs* are directly or indirectly induced by the downstream components of the Wnt-1 signaling pathway (i.e., β-catenin-TCF-1/Lef1). The increased levels of *WISP* RNA were measured in Wnt-1-transformed cells, hours or days after Wnt-1 transformation. Thus, *WISP* expression could result from Wnt-1 signaling directly through β-catenin transcription factor regulation or alternatively through Wnt-1 signaling turning on a transcription factor, which in turn regulates *WISPs*.

The *WISPs* define an additional subfamily of the CCN family of growth factors. One striking difference observed in the protein sequence of WISP-2 is the absence of a CT domain, which is present in CTGF, Cyr61, *nov*, WISP-1, and WISP-3. This domain is thought to be involved in receptor binding and dimerization. Growth factors, such as TGF-β, platelet-derived growth factor, and nerve growth factor, which contain a cystine knot motif exist as dimers (32). It is tempting to speculate that WISP-1 and WISP-3 may exist as dimers, whereas WISP-2 exists as a monomer. If the CT domain is also important for receptor binding, WISP-2 may bind its receptor through a different region of the molecule than the other CCN family members. No specific receptors have been identified for CTGF or *nov*. A recent report has shown that integrin $\alpha_v\beta_3$ serves as an adhesion receptor for Cyr61 (33).

The strong expression of *WISP-1* and *WISP-2* in cells lying within the fibrovascular tumor stroma in breast tumors from Wnt-1 transgenic animals is consistent with previous observations that transcripts for the related CTGF gene are primarily expressed in the fibrous stroma of mammary tumors (34). Epithelial cells are thought to control the proliferation of connective tissue stroma in mammary tumors by a cascade of growth factor signals similar to that controlling connective tissue formation during wound repair. It has been proposed that mammary tumor cells or inflammatory cells at the tumor interstitial interface secrete TGF-β1, which is the stimulus for stromal proliferation (34). TGF-β1 is secreted by a large percentage of malignant breast tumors and may be one of the growth factors that stimulates the production of CTGF and WISPs in the stroma.

It was of interest that *WISP-1* and *WISP-2* expression was observed in the stromal cells that surrounded the tumor cells

(epithelial cells) in the Wnt-1 transgenic mouse sections of breast tissue. This finding suggests that paracrine signaling could occur in which the stromal cells could supply WISP-1 and WISP-2 to regulate tumor cell growth on the WISP extracellular matrix. Stromal cell-derived factors in the extracellular matrix have been postulated to play a role in tumor cell migration and proliferation (35). The localization of *WISP-1* and *WISP-2* in the stromal cells of breast tumors supports this paracrine model.

An analysis of *WISP-1* gene amplification and expression in human colon tumors showed a correlation between DNA amplification and overexpression, whereas overexpression of *WISP-3* RNA was seen in the absence of DNA amplification. In contrast, *WISP-2* DNA was amplified in the colon tumors, but its mRNA expression was significantly reduced in the majority of tumors compared with the expression in normal colonic mucosa from the same patient. The gene for human *WISP-2* was localized to chromosome 20q12–20q13, at a region frequently amplified and associated with poor prognosis in node negative breast cancer and many colon cancers, suggesting the existence of one or more oncogenes at this locus (36–38). Because the center of the 20q13 amplicon has not yet been identified, it is possible that the apparent amplification observed for *WISP-2* may be caused by another gene in this amplicon.

A recent manuscript on *rCop-1*, the rat orthologue of *WISP-2*, describes the loss of expression of this gene after cell transformation, suggesting it may be a negative regulator of growth in cell lines (16). Although the mechanism by which *WISP-2* RNA expression is down-regulated during malignant transformation is unknown, the reduced expression of *WISP-2* in colon tumors and cell lines suggests that it may function as a tumor suppressor. These results show that the *WISP* genes are aberrantly expressed in colon cancer and suggest that their altered expression may confer selective growth advantage to the tumor.

Members of the Wnt signaling pathway have been implicated in the pathogenesis of colon cancer, breast cancer, and melanoma, including the tumor suppressor gene adenomatous polyposis coli and β-catenin (39). Mutations in specific regions of either gene can cause the stabilization and accumulation of cytoplasmic β-catenin, which presumably contributes to human carcinogenesis through the activation of target genes such as the *WISPs*. Although the mechanism by which Wnt-1 transforms cells and induces tumorigenesis is unknown, the identification of *WISPs* as genes that may be regulated downstream of Wnt-1 in CS7MG cells suggests they could be important mediators of Wnt-1 transformation. The amplification and altered expression patterns of the *WISPs* in human colon tumors may indicate an important role for these genes in tumor development.

We thank the DNA synthesis group for oligonucleotide synthesis, T. Baker for technical assistance, F. Dowd for radiation hybrid mapping, K. Willert and R. Nusse for the tet-repressible CS7MG/Wnt-1 cells, V. Dixit for discussions, and D. Wood and A. Bruce for artwork.

1. Cadigan, K. M. & Nusse, R. (1997) *Genes Dev.* 11, 3286–3305.
2. Dale, T. C. (1998) *Biochem. J.* 329, 209–223.
3. Nusse, R. & Varmus, H. E. (1982) *Cell* 31, 99–109.
4. van Ooyen, A. & Nusse, R. (1984) *Cell* 39, 233–240.
5. Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. (1988) *Cell* 55, 619–625.
6. Brown, J. D. & Moon, R. T. (1998) *Curr. Opin. Cell. Biol.* 10, 182–187.
7. Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. & Clevers, H. (1996) *Cell* 86, 391–399.
8. Korinek, V., Barker, N., Willert, K., Molenaar, M., Roose, J., Wagenaar, G., Markman, M., Lamers, W., Destree, O. & Clevers, H. (1998) *Mol. Cell. Biol.* 18, 1248–1256.
9. Matsumoto, S., Albert, I., Souza, B., Rubinfeld, B. & Polakis, P. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3046–3050.
10. He, T. C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B. & Kinzler, K. W. (1998) *Science* 281, 1509–1512.
11. Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 6025–6030.
12. Brown, A. M., Wildin, R. S., Prendergast, T. J. & Varmus, H. E. (1986) *Cell* 46, 1001–1009.
13. Wong, G. T., Gavin, B. J. & McMahon, A. P. (1994) *Mol. Cell. Biol.* 14, 6278–6286.
14. Shimizu, H., Julius, M. A., Giarre, M., Zheng, Z., Brown, A. M. & Kitajewski, J. (1997) *Cell Growth Differ.* 8, 1349–1358.
15. Hashimoto, Y., Shindo-Okada, N., Tani, M., Nagamachi, Y., Takeuchi, K., Shiroishi, T., Toma, H. & Yokota, J. (1998) *J. Exp. Med.* 187, 289–296.
16. Zhang, R., Averboukh, L., Zhu, W., Zhang, H., Jo, H., Dempsey, P. J., Coffey, R. J., Pardee, A. B. & Liang, P. (1998) *Mol. Cell. Biol.* 18, 6131–6141.
17. Grotendorst, G. R. (1997) *Cytokine Growth Factor Rev.* 8, 171–179.
18. Kireeva, M. L., Mo, F. E., Yang, G. P. & Lau, L. F. (1996) *Mol Cell. Biol.* 16, 1326–1334.
19. Babic, A. M., Kireeva, M. L., Kolesnikova, T. V. & Lau, L. F. (1998) *Proc. Natl. Acad. Sci. USA* 95, 6355–6360.
20. Martinerie, C., Huff, V., Joubert, I., Hadzioch, M., Saunders, G., Strong, L. & Perbal, B. (1994) *Oncogene* 9, 2729–2732.
21. Bork, P. (1993) *FEBS Lett.* 327, 125–130.
22. Kim, H. S., Nagalla, S. R., Oh, Y., Wilson, E., Roberts, C. T., Jr. & Rosenfeld, R. G. (1997) *Proc. Natl. Acad. Sci. USA* 94, 12981–12986.
23. Joliot, V., Martinerie, C., Dambrine, G., Plassiart, G., Brisac, M., Crochet, J. & Perbal, B. (1992) *Mol. Cell. Biol.* 12, 10–21.
24. Mancuso, D. J., Tuley, E. A., Westfield, L. A., Worrall, N. K., Shelton-Inloes, B. B., Sorace, J. M., Alevy, Y. G. & Sadler, J. E. (1989) *J. Biol. Chem.* 264, 19514–19527.
25. Holt, O. D., Pangburn, M. K. & Ginsburg, V. (1990) *J. Biol. Chem.* 265, 2852–2855.
26. Voorberg, J., Fontijn, R., Calafat, J., Janssen, H., van Mourik, J. A. & Pannekoek, H. (1991) *J. Cell. Biol.* 113, 195–205.
27. Martinerie, C., Viegas-Pequignot, E., Guenard, I., Dutrillaux, B., Nguyen, V. C., Bernheim, A. & Perbal, B. (1992) *Oncogene* 7, 2529–2534.
28. Takahashi, E., Hori, T., O'Connell, P., Leppert, M. & White, R. (1991) *Cytogenet. Cell. Genet.* 57, 109–111.
29. Meese, E., Meltzer, P. S., Witkowski, C. M. & Trent, J. M. (1989) *Genes Chromosomes Cancer* 1, 88–94.
30. Carter, S. J. (1994) *Crit. Rev. Oncog.* 4, 435–449.
31. Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 276, 1268–1272.
32. Sun, P. D. & Davies, D. R. (1995) *Annu. Rev. Biophys. Biomol. Struct.* 24, 269–291.
33. Kireeva, M. L., Lam, S. C. T. & Lau, L. F. (1998) *J. Biol. Chem.* 273, 3090–3096.
34. Frazier, K. S. & Grotendorst, G. R. (1997) *Int. J. Biochem. Cell Biol.* 29, 153–161.
35. Werner, N. (1997) *Virchows Arch.* 430, 433–443.
36. Tanner, M. M., Tirkkonen, M., Kallioniemi, A., Collins, C., Stokka, T., Karhu, R., Kowbel, D., Shadravan, F., Hintz, M., Kuo, W. L., *et al.* (1994) *Cancer Res.* 54, 4257–4260.
37. Brinkmann, U., Gallo, M., Polymeropoulos, M. H. & Pastan, I. (1996) *Genome Res.* 6, 187–194.
38. Bieche, I., Tozlu, S., Vidaud, D., Laurendeau, I., Lidereau, R., Vidaud, M. & Champeme, M. H. (1998) *Int. J. Oncol.* 12, 1155–1160.
39. Morin, P. J., Sparks, A. B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B. & Kinzler, K. W. (1997) *Science* 275, 1787–1790.
40. Lu, L. H. & Gillett, N. (1994) *Cell Vision* 1, 160–176.

HellerEhrman
ATTORNEYS

GENOME METHODS

Real Time Quantitative PCR

Christian A. Heid,[1] Junko Stevens,[2] Kenneth J. Livak,[2] and P. Mickey Williams[1,3]

[1]BioAnalytical Technology Department, Genentech, Inc., South San Francisco, California 94080; [2]Applied BioSystems Division of Perkin Elmer Corp., Foster City, California 94404

We have developed a novel "real time" quantitative PCR method. The method measures PCR product accumulation through a dual-labeled fluorogenic probe (i.e., TaqMan Probe). This method provides very accurate and reproducible quantitation of gene copies. Unlike other quantitative PCR methods, real-time PCR does not require post-PCR sample handling, preventing potential PCR product carry-over contamination and resulting in much faster and higher throughput assays. The real-time PCR method has a very large dynamic range of starting target molecule determination (at least five orders of magnitude). Real-time quantitative PCR is extremely accurate and less labor-intensive than current quantitative PCR methods.

Quantitative nucleic acid sequence analysis has had an important role in many fields of biological research. Measurement of gene expression (RNA) has been used extensively in monitoring biological responses to various stimuli (Tan et al. 1994; Huang et al. 1995a,b; Prud'homme et al. 1995). Quantitative gene analysis (DNA) has been used to determine the genomic quantity of a particular gene, as in the case of the human *HER2* gene, which is amplified in ~30% of breast tumors (Slamon et al. 1987). Gene and genome quantitation (DNA and RNA) also have been used for analysis of human immunodeficiency virus (HIV) burden demonstrating changes in the levels of virus throughout the different phases of the disease (Connor et al. 1993; Piatak et al. 1993b; Furtado et al. 1995).

Many methods have been described for the quantitative analysis of nucleic acid sequences (both for RNA and DNA; Southern 1975; Sharp et al. 1980; Thomas 1980). Recently, PCR has proven to be a powerful tool for quantitative nucleic acid analysis. PCR and reverse transcriptase (RT)–PCR have permitted the analysis of minimal starting quantities of nucleic acid (as little as one cell equivalent). This has made possible many experiments that could not have been performed with traditional methods. Although PCR has provided a powerful tool, it is imperative that it be used properly for quantitation (Raeymaekers 1995). Many early reports of quantitative PCR and RT–PCR described quantitation of the PCR product but did not measure the initial target sequence quantity. It is essential to design proper controls for the quantitation of the initial target sequences (Ferre 1992; Clementi et al. 1993).

Researchers have developed several methods of quantitative PCR and RT–PCR. One approach measures PCR product quantity in the log phase of the reaction before the plateau (Kellogg et al. 1990; Pang et al. 1990). This method requires that each sample has equal input amounts of nucleic acid and that each sample under analysis amplifies with identical efficiency up to the point of quantitative analysis. A gene sequence (contained in all samples at relatively constant quantities, such as β-actin) can be used for sample amplification efficiency normalization. Using conventional methods of PCR detection and quantitation (gel electrophoresis or plate capture hybridization), it is extremely laborious to assure that all samples are analyzed during the log phase of the reaction (for both the target gene and the normalization gene). Another method, quantitative competitive (QC)–PCR, has been developed and is used widely for PCR quantitation. QC–PCR relies on the inclusion of an internal control competitor in each reaction (Becker-Andre 1991; Piatak et al. 1993a,b). The efficiency of each reaction is normalized to the internal competitor. A known amount of internal competitor can be

[3]Corresponding author.

added to each sample. To obtain relative quantitation, the unknown target PCR product is compared with the known competitor PCR product. Success of a quantitative competitive PCR assay relies on developing an internal control that amplifies with the same efficiency as the target molecule. The design of the competitor and the validation of amplification efficiencies require a dedicated effort. However, because QC–PCR does not require that PCR products be analyzed during the log phase of the amplification, it is the easier of the two methods to use.

Several detection systems are used for quantitative PCR and RT-PCR analysis: (1) agarose gels, (2) fluorescent labeling of PCR products and detection with laser-induced fluorescence using capillary electrophoresis (Fuoco et al. 1995; Williams et al. 1996) or acrylamide gels, and (3) plate capture and sandwich probe hybridization (Mulder et al. 1994). Although these methods proved successful, each method requires post-PCR manipulations that add time to the analysis and may lead to laboratory contamination. The sample throughput of these methods is limited (with the exception of the plate capture approach), and, therefore, these methods are not well suited for uses demanding high sample throughput (i.e., screening of large numbers of biomolecules or analyzing samples for diagnostics or clinical trials).

Here we report the development of a novel assay for quantitative DNA analysis. The assay is based on the use of the 5′ nuclease assay first described by Holland et al. (1991). The method uses the 5′ nuclease activity of *Taq* polymerase to cleave a nonextendible hybridization probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee et al. 1993; Bassler et al. 1995; Livak et al. 1995a,b). One fluorescent dye serves as a reporter [FAM (i.e., 6-carboxyfluorescein)] and its emission spectra is quenched by the second fluorescent dye, TAMRA (i.e., 6-carboxy-tetramethylrhodamine). The nuclease degradation of the hybridization probe releases the quenching of the FAM fluorescent emission, resulting in an increase in peak fluorescent emission at 518 nm. The use of a sequence detector (ABI Prism) allows measurement of fluorescent spectra of all 96 wells of the thermal cycler continuously during the PCR amplification. Therefore, the reactions are monitored in real time. The output data is described and quantitative analysis of input target DNA sequences is discussed below.

RESULTS

PCR Product Detection in Real Time

The goal was to develop a high-throughput, sensitive, and accurate gene quantitation assay for use in monitoring lipid mediated therapeutic gene delivery. A plasmid encoding human factor VIII gene sequence, pBSTM (see Methods), was used as a model therapeutic gene. The assay uses fluorescent Taqman methodology and an instrument capable of measuring fluorescence in real time (ABI Prism 7700 Sequence Detector). The Taqman reaction requires a hybridization probe labeled with two different fluorescent dyes. One dye is a reporter dye (FAM), the other is a quenching dye (TAMRA). When the probe is intact fluorescent energy transfer occurs and the reporter dye fluorescent emission is absorbed by the quenching dye (TAMRA). During the extension phase of the PCR cycle, the fluorescent hybridization probe is cleaved by the 5′–3′ nucleolytic activity of the DNA polymerase. On cleavage of the probe, the reporter dye emission is no longer transferred efficiently to the quenching dye, resulting in an increase of the reporter dye fluorescent emission spectra. PCR primers and probes were designed for the human factor VIII sequence and human β-actin gene (as described in Methods). Optimization reactions were performed to choose the appropriate probe and magnesium concentrations yielding the highest intensity of reporter fluorescent signal without sacrificing specificity. The instrument uses a charge-coupled device (i.e., CCD camera) for measuring the fluorescent emission spectra from 500 to 650 nm. Each PCR tube was monitored sequentially for 25 msec with continuous monitoring throughout the amplification. Each tube was re-examined every 8.5 sec. Computer software was designed to examine the fluorescent intensity of both the reporter dye (FAM) and the quenching dye (TAMRA). The fluorescent intensity of the quenching dye, TAMRA, changes very little over the course of the PCR amplification (data not shown). Therefore, the intensity of TAMRA dye emission serves as an internal standard with which to normalize the reporter dye (FAM) emission variations. The software calculates a value termed ΔRn (or ΔRQ) using the following equation: $\Delta Rn = (Rn^+) - (Rn^-)$, where Rn^+ = emission intensity of reporter/emission intensity of quencher at any given time in a reaction tube, and Rn^- = emission intensity of re-

porter/emission intensity of quencher measured prior to PCR amplification in that same reaction tube. For the purpose of quantitation, the last three data points (ΔRns) collected during the extension step for each PCR cycle were analyzed. The nucleolytic degradation of the hybridization probe occurs during the extension phase of PCR, and, therefore, reporter fluorescent emission increases during this time. The three data points were averaged for each PCR cycle and the mean value for each was plotted in an "amplification plot" shown in Figure 1A. The ΔRn mean value is plotted on the *y*-axis, and time, represented by cycle number, is plotted on the *x*-axis. During the early cycles of the PCR amplification, the ΔRn value remains at base line. When sufficient hybridization probe has been cleaved by the *Taq* polymerase nuclease activity, the intensity of reporter fluorescent emission increases. Most PCR amplifications reach a plateau phase of reporter fluorescent emission if the reaction is carried out to high cycle numbers. The amplification plot is examined early in the reaction, at a point that represents the log phase of product accumulation. This is done by assigning an arbitrary threshold that is based on the variability of the base-line data. In Figure 1A, the threshold was set at 10 standard deviations above the mean of base line emission calculated from cycles 1 to 15. Once the threshold is chosen, the point at which

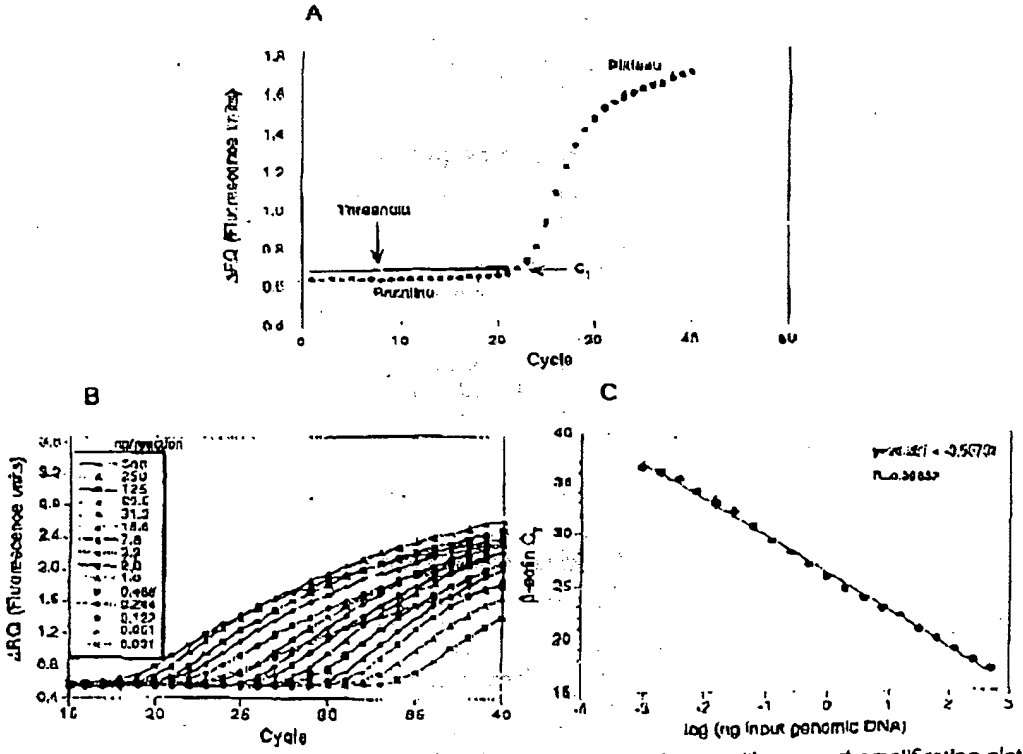


Figure 1 PCR product detection in real time. (A) The Model 7700 software will construct amplification plots from the extension phase fluorescent emission data collected during the PCR amplification. The standard deviation is determined from the data points collected from the base line of the amplification plot. C_T values are calculated by determining the point at which the fluorescence exceeds a threshold limit (usually 10 times the standard deviation of the base line). (B) Overlay of amplification plots of serially (1:2) diluted human genomic DNA samples amplified with β-actin primers. (C) Input DNA concentration of the samples plotted versus C_T. All

the amplification plot crosses the threshold is defined as C_T. C_T is reported as the cycle number at this point. As will be demonstrated, the C_T value is predictive of the quantity of input target.

C_T Values Provide a Quantitative Measurement of Input Target Sequences

Figure 1B shows amplification plots of 15 different PCR amplifications overlaid. The amplifications were performed on a 1:2 serial dilution of human genomic DNA. The amplified target was human β actin. The amplification plots shift to the right (to higher threshold cycles) as the input target quantity is reduced. This is expected because reactions with fewer starting copies of the target molecule require greater amplification to degrade enough probe to attain the threshold fluorescence. An arbitrary threshold of 10 standard deviations above the base line was used to determine the C_T values. Figure 1C represents the C_T values plotted versus the sample dilution value. Each dilution was amplified in triplicate PCR amplifications and plotted as mean values with error bars representing one standard deviation. The C_T values decrease linearly with increasing target quantity. Thus, C_T values can be used as a quantitative measurement of the input target number. It should be noted that the amplification plot for the 15.6-ng sample shown in Figure 1B does not reflect the same fluorescent rate of increase exhibited by most of the other samples. The 15.6-ng sample also achieves endpoint plateau at a lower fluorescent value than would be expected based on the input DNA. This phenomenon has been observed occasionally with other samples (data not shown) and may be attributable to late cycle inhibition; this hypothesis is still under investigation. It is important to note that the flattened slope and early plateau do not impact significantly the calculated C_T value as demonstrated by the fit on the line shown in Figure 1C. All triplicate amplifications resulted in very similar C_T values—the standard deviation did not exceed 0.5 for any dilution. This experiment contains a >100,000-fold range of input target molecules. Using C_T values for quantitation permits a much larger assay range than directly using total fluorescent emission intensity for quantitation. The linear range of fluorescent intensity measurement of the ABI Prism 7700 Se- ments over a very large range of relative starting target quantities.

Sample Preparation Validation

Several parameters influence the efficiency of PCR amplification: magnesium and salt concentrations, reaction conditions (i.e., time and temperature), PCR target size and composition, primer sequences, and sample purity. All of the above factors are common to a single PCR assay, except sample to sample purity. In an effort to validate the method of sample preparation for the factor VIII assay, PCR amplification reproducibility and efficiency of 10 replicate sample preparations were examined. After genomic DNA was prepared from the 10 replicate samples, the DNA was quantitated by ultraviolet spectroscopy. Amplifications were performed analyzing β-actin gene content in 100 and 25 ng of total genomic DNA. Each PCR amplification was performed in triplicate. Comparison of C_T values for each triplicate sample show minimal variation based on standard deviation and coefficient of variance (Table 1). Therefore, each of the triplicate PCR amplifications was highly reproducible, demonstrating that real time PCR using this instrumentation introduces minimal variation into the quantitative PCR analysis. Comparison of the mean C_T values of the 10 replicate sample preparations also showed minimal variability, indicating that each sample preparation yielded similar results for β-actin gene quantity. The highest C_T difference between any of the samples was 0.85 and 0.73 for the 100 and 25 ng samples, respectively. Additionally, the amplification of each sample exhibited an equivalent rate of fluorescent emission intensity change per amount of DNA target analyzed as indicated by similar slopes derived from the sample dilutions (Fig. 2). Any sample containing an excess of a PCR inhibitor would exhibit a greater measured β-actin C_T value for a given quantity of DNA. In addition, the inhibitor would be diluted along with the sample in the dilution analysis (Fig. 2), altering the expected C_T value change. Each sample amplification yielded a similar result in the analysis, demonstrating that this method of sample preparation is highly reproducible with regard to sample purity.

Quantitative Analysis of a Plasmid After

MID-PAL

Table 1. Reproducibility of Sample Preparation Method

	100 ng				25 ng			
Sample no.	C_T	mean	standard deviation	CV	C_T	mean	standard deviation	CV
1	18.24				20.48			
	18.23				20.55			
	18.33	18.27	0.06	0.32	20.5	20.51	0.03	0.17
2	18.33				20.61			
	18.35				20.59			
	18.44	18.37	0.06	0.32	20.41	20.54	0.11	0.54
3	18.3				20.54			
	18.3				20.6			
	18.42	18.34	0.07	0.36	20.49	20.54	0.06	0.28
4	18.15				20.48			
	18.23				20.44			
	18.32	18.23	0.08	0.46	20.38	20.43	0.05	0.26
5	18.4				20.68			
	18.38				20.87			
	18.46	18.42	0.04	0.23	20.63	20.71	0.13	0.61
6	18.54				21.09			
	18.67				21.04			
	19	18.74	0.24	1.26	21.01	21.06	0.03	0.15
7	18.28				20.67			
	18.36				20.73			
	18.57	18.39	0.12	0.66	20.65	20.68	0.04	0.2
8	18.45				20.98			
	18.7				20.84			
	18.73	18.63	0.16	0.83	20.75	20.86	0.12	0.57
9	18.18				20.46			
	18.34				20.54			
	18.36	18.29	0.1	0.55	20.48	20.51	0.07	0.32
10	18.42				20.79			
	18.57				20.78			
	18.66	18.55	0.12	0.66	20.62	20.73	0.1	0.46
Mean	(1-10)	18.42	0.17	0.90		20.66	0.19	0.94

for containing a partial cDNA for human factor VIII, pF8TM. A series of transfections was set up using a decreasing amount of the plasmid (40, 4, 0.5, and 0.1 μg). Twenty-four hours post-transfection, total DNA was purified from each flask of cells. β-Actin gene quantity was chosen as a value for normalization of genomic DNA concentration from each sample. In this experiment, β-actin gene content should remain constant relative to total genomic DNA. Figure 3 shows the result of the β-actin DNA measurement (100 ng total DNA determined by ultraviolet spectroscopy) of each sample. Each sample was analyzed in triplicate and the mean β-actin C_T values of the triplicates were plotted (error bars represent one standard deviation). The highest difference between any two sample means was 0.95 C_T. Ten nanograms of total DNA of each sample were also examined for β-actin. The results again showed that very similar amounts of genomic DNA were present; the maximum mean β-actin C_T value difference was 1.0. As Figure 3 shows, the rate of β-actin C_T change between the 100- and 10-ng samples was similar (slope values range between −3.56 and −3.45). This verifies again that the method of sample preparation yields samples of identical PCR integrity (i.e., no sample contained an excessive amount of a PCR inhibitor). However, these results indicate that each sample contained slight differences in the actual amount of genomic DNA analyzed. Determination of actual genomic DNA concentration was accomplished

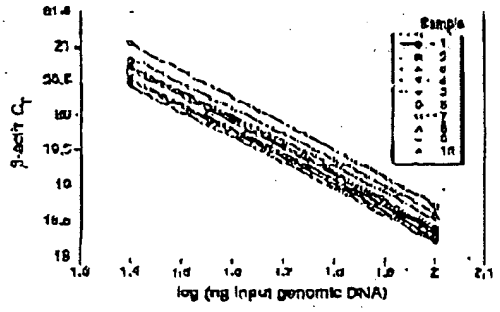


Figure 2 Sample preparation purity. The replicate samples shown in Table 1 were also amplified in triplicate using 25 ng of each DNA sample. The figure shows the input DNA concentration (100 and 25 ng) vs. C_T. In the figure, the 100 and 25 ng points for each sample are connected by a line.

by plotting the mean β-actin C_T value obtained for each 100-ng sample on a β-actin standard curve (shown in Fig. 4C). The actual genomic DNA concentration of each sample, *a*, was obtained by extrapolation to the *x*-axis.

Figure 4A shows the measured (i.e., uncorrected) quantities of factor VIII plasmid DNA (pF8TM) from each of the four transient cell transfections. Each reaction contained 100 ng of total sample DNA (as determined by UV spectroscopy). Each sample was analyzed in triplicate

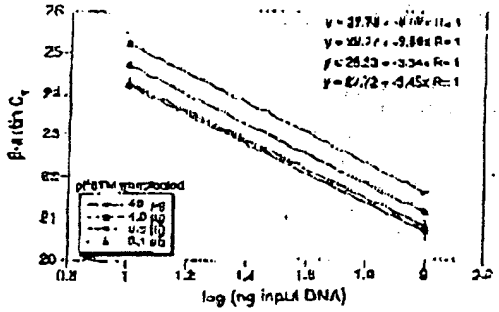


Figure 3 Analysis of transfected cell DNA quantity and purity. The DNA preparations of the four 293 cell transfections (40, 4, 0.5, and 0.1 μg of pF8TM) were analyzed for the β-actin gene. 100 and 10 ng (determined by ultraviolet spectroscopy) of each sample were amplified in triplicate. For each amount of pF8TM that was transfected, the β-actin C_T values are plotted versus the total input DNA concentration.

PCR amplifications. As shown, pF8TM purified from the 293 cells decreases (mean C_T values increase) with decreasing amounts of plasmid transfected. The mean C_T values obtained for pF8TM in Figure 4A were plotted on a standard curve comprised of serially diluted pF8TM, shown in Figure 4B. The quantity of pF8TM, *b*, found in each of the four transfections was determined by extrapolation to the *x* axis of the standard curve in Figure 4B. These uncorrected values, *b*, for pF8TM were normalized to determine the actual amount of pF8TM found per 100 ng of genomic DNA by using the equation:

$$\frac{b \times 100\ \text{ng}}{a} = \frac{\text{actual pF8TM copies per}}{\text{100 ng of genomic DNA}}$$

where *a* = actual genomic DNA in a sample and *b* = pF8TM copies from the standard curve. The normalized quantity of pF8TM per 100 ng of genomic DNA for each of the four transfections is shown in Figure 4D. These results show that the quantity of factor VIII plasmid associated with the 293 cells, 24 hr after transfection, decreases with decreasing plasmid concentration used in the transfection. The quantity of pF8TM associated with 293 cells, after transfection with 40 μg of plasmid, was 35 pg per 100 ng genomic DNA. This results in ~520 plasmid copies per cell.

DISCUSSION

We have described a new method for quantitating gene copy numbers using real-time analysis of PCR amplifications. Real-time PCR is compatible with either of the two PCR (RT–PCR) approaches: (1) quantitative competitive where an internal competitor for each target sequence is used for normalization (data not shown) or (2) quantitative comparative PCR using a normalization gene contained within the sample (i.e., β-actin) or a "housekeeping" gene for RT–PCR. If equal amounts of nucleic acid are analyzed for each sample and if the amplification efficiency before quantitative analysis is identical for each sample, the internal control (normalization gene or competitor) should give equal signals for all samples.

The real-time PCR method offers several advantages over the other two methods currently employed (see the Introduction). First, the real-time PCR method is performed in a closed-tube system and requires no post-PCR manipulation

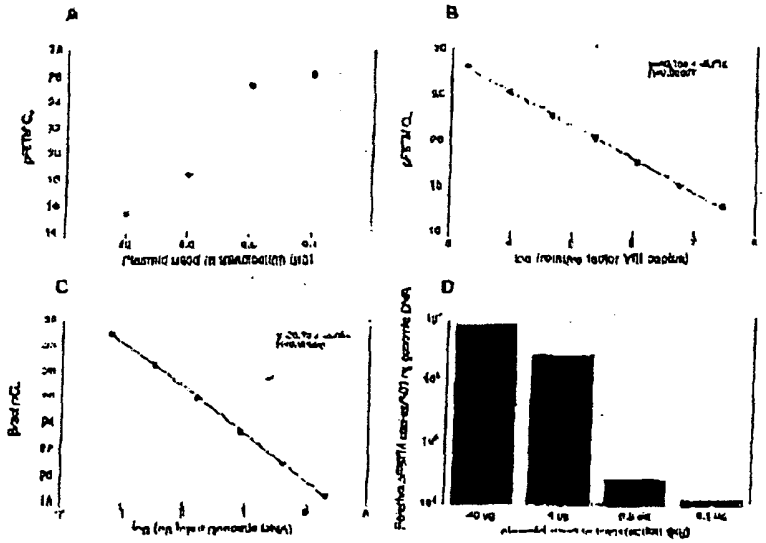


Figure 4 Quantitative analysis of pF8TM in transfected cells. (*A*) Amount of plasmid DNA used for the transfection plotted against the mean C_T value determined for pF8TM remaining 24 hr after transfection. (*B*,*C*) Standard curves of pF8TM and β-actin, respectively. pF8TM DNA (*B*) and genomic DNA (*C*) were diluted serially 1:5 before amplification with the appropriate primers. The β-actin standard curve was used to normalize the results of *A* to 100 ng of genomic DNA. (*D*) The amount of pF8TM present per 100 ng of genomic DNA.

of sample. Therefore, the potential for PCR contamination in the laboratory is reduced because amplified products can be analyzed and disposed of without opening the reaction tubes. Second, this method supports the use of a normalization gene (i.e., β-actin) for quantitative PCR or housekeeping genes for quantitative RT-PCR controls. Analysis is performed in real time during the log phase of product accumulation. Analysis during log phase permits many different genes (over a wide input target range) to be analyzed simultaneously, without concern of reaching reaction plateau at different cycles. This will make multigene analysis assays much easier to develop, because individual internal competitors will not be needed for each gene under analysis. Third, sample throughput will increase dramatically with the new method because there is no post-PCR processing time. Additionally, working in a 96-well format is highly compatible with automation technology.

The real-time PCR method is highly reproducible. Replicate amplifications can be analyzed for each sample minimizing potential error. The system allows for a very large assay dynamic range (approaching 1,000,000-fold starting target). Using a standard curve for the target of interest, relative copy number values can be determined for any unknown sample. Fluorescent threshold values, C_T, correlate linearly with relative DNA copy numbers. Real time quantitative RT-PCR methodology (Gibson et al., this issue) has also been developed. Finally, real time quantitative PCR methodology can be used to develop high-throughput screening assays for a variety of applications (quantitative gene expression (RT-PCR), gene copy assays (Her2, HIV, etc.), genotyping (knockout mouse analysis), and Immuno-PCR).

Real-time PCR may also be performed using intercalating dyes (Higuchi et al. 1992) such as ethidium bromide. The fluorogenic probe method offers a major advantage over intercalating dyes—greater specificity (i.e., primer dimers and nonspecific PCR products are not detected).

METHODS

Generation of a Plasmid Containing a Partial cDNA for Human Factor VIII

Total RNA was harvested (RNAzol B from Tel-Test, Inc., Friendswood, TX) from cells transfected with a factor VIII expression vector, pCIS2.8c25D (Eaton et al. 1986; Gorman et al. 1990). A factor VIII partial cDNA sequence was generated by RT-PCR (GeneAmp EZ rTth RNA PCR Kit (part N808-0179, PE Applied Biosystems, Foster City, CA)) using the PCR primers F8for and F8rev (primer sequences are shown below). The amplicon was reamplified using modified F8for and F8rev primers (appended with BamHI and HindIII restriction site sequences at the 5' end) and cloned into pGEM-3Z (Promega Corp., Madison, WI). The resulting clone, pF8TM, was used for transient transfection of 293 cells.

Amplification of Target DNA and Detection of Amplicon Factor VIII Plasmid DNA

(pF8TM) was amplified with the primers F8for 5'-CCCGTGCCAAGAGTGACTTCTTC-3' and F8rev 5'-AAACCTCACCCTGGATGGTTAGG-3'. The reaction produced a 422-bp PCR product. The forward primer was designed to recognize a unique sequence found in the 5' untranslated region of the parent pCIS2.8c25D plasmid and therefore does not recognize and amplify the human factor VIII gene. Primers were chosen with the assistance of the computer program Oligo 4.0 (National Biosciences, Inc., Plymouth, MN). The human β-actin gene was amplified with the primers β-actin forward primer 5'-TCACCCACACTGTGCCCATCTACGA-3' and β-actin reverse primer 5'-CAGCGGAACCGCTCATTGCCAATGG-3'. The reaction produced a 295-bp PCR product.

Amplification reactions (50 µl) contained a DNA sample, 10× PCR Buffer II (5 µl), 200 µM dATP, dCTP, dGTP, and 400 µM dUTP, 4 mM $MgCl_2$, 1.25 Units AmpliTaq DNA polymerase, 0.5 unit AmpErase uracil N-glycosylase (UNG), 50 pmole of each factor VIII primer, and 15 pmole of each β-actin primer. The reactions also contained one of the following detection probes (100 nM each): F8probe 5'(FAM)AGCTCTCCACCTGCTTCTTTCTGTGCCTT(TAMRA)p-3' and β-actin probe 5'(FAM)ATGCCC-X(TAMRA)CCCCCATGCCATCp-3' where p indicates phosphorylation and X indicates a linker arm nucleotide. Reaction tubes were MicroAmp Optical Tubes (part number N801-0933, Perkin Elmer) that were frosted (at Perkin Elmer) to prevent light from reflecting. Tube caps were similar to MicroAmp Caps but specially designed to prevent light scattering. All of the PCR consumables were supplied by PE Applied Biosystems (Foster City, CA) except the factor VIII primers, which were synthesized at Genentech, Inc. (South San Francisco, CA). Probes were designed using the Oligo 4.0 software, following guidelines suggested in the Model 7700 Sequence Detector instrument manual. Briefly, probe T_m should be at least 5°C higher than the annealing temperature used during thermal cycling; primers should not form stable duplexes with the probe.

The thermal cycling conditions included 2 min at 50°C and 10 min at 95°C. Thermal cycling proceeded with reactions were performed in the Model 7700 Sequence Detector (PE Applied Biosystems), which contains a GeneAmp PCR System 9600. Reaction conditions were programmed on a Power Macintosh 7100 (Apple Computer, Santa Clara, CA) linked directly to the Model 7700 Sequence Detector. Analysis of data was also performed on the Macintosh computer. Collection and analysis software was developed at PE Applied Biosystems.

Transfection of Cells with Factor VIII Construct

Four T175 flasks of 293 cells (ATCC CRL 1573), a human fetal kidney suspension cell line, were grown to 80% confluency and transfected pF8TM. Cells were grown in the following media: 50% HAM'S F12 without GHT, 50% low glucose Dulbecco's modified Eagle medium (DMEM) without glycine with sodium bicarbonate, 10% fetal bovine serum, 2 mM L-glutamine, and 1% penicillin-streptomycin. The media was changed 30 min before the transfection. pF8TM DNA amounts of 40, 4, 0.5, and 0.1 µg were added to 1.6 ml of a solution containing 0.125 M $CaCl_2$ and 1× HBPS. The four mixtures were left at room temperature for 10 min and then added dropwise to the cells. The flasks were incubated at 37°C and 5% CO_2 for 24 hr, washed with PBS, and resuspended in PBS. The resuspended cells were divided into aliquots and DNA was extracted immediately using the QIAamp Blood Kit (Qiagen, Chatsworth, CA). DNA was eluted into 200 µl of 20 mM Tris-HCl at pH 8.0.

ACKNOWLEDGMENTS

We thank Genentech's DNA Synthesis Group for primer synthesis and Genentech's Graphics Group for assistance with the figures.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Bassler, H.A., S.J. Flood, K.J. Livak, J. Marmaro, R. Knorr, and C.A. Batt. 1995. Use of a fluorogenic probe in a PCR-based assay for the detection of Listeria monocytogenes. *App. Environ. Microbiol.* 61: 3724–3728.

Becker-Andre, M. 1991. Quantitative evaluation of mRNA levels. *Meth. Mol. Cell. Biol.* 2: 189–201.

Clementi, M., S. Menzo, P. Bagnarelli, A. Manzin, A. Valenza, and P.E. Varaldo. 1993. Quantitative PCR and RT-PCR in virology. [Review]. *PCR Methods Applic.* 2: 191–196.

Connor, R.I., H. Mohri, Y. Cao, and D.D. Ho. 1993. Increased viral burden and cytopathicity correlate temporally with CD4+ T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 67: 1772–1777.

Eaton, D.L., W.I. Wood, D. Eaton, P.E. Hass, P.

Vehar, and G. Gorman. 1986. Construction and characterization of an active factor VIII variant lacking the central one third of the molecule. *Biochemistry* 25: 8343–8347.

Fasco, M.J., C.P. Treanor, S. Spivack, H.L. Figge, and L.S. Kaminsky. 1995. Quantitative RNA-polymerase chain reaction-DNA analysis by capillary electrophoresis and laser-induced fluorescence. *Anal. Biochem.* 224: 140–147.

Forte, R. 1992. Quantitative or semi-quantitative PCR: Reality versus myth. *PCR Methods Applic.* 2: 1–9.

Furtado, M.R., L.A. Kingsley, and S.M. Wolinsky. 1994. Changes in the viral mRNA expression pattern correlate with a rapid rate of CD4+ T-cell number decline in human immunodeficiency virus type 1-infected individuals. *J. Virol.* 68: 2092–2100.

Gibson, U.E.M., C.A. Heid, and P.M. Williams. 1996. A novel method for real time quantitative competitive RT-PCR. *Genome Res.* (this issue).

Guttman, C.M., D.R. Glas, and G. McCray. 1990. Transient production of proteins using an adenovirus transformed cell line. *DNA Prot. Engin. Tech.* 2: 3–10.

Higuchi, R., G. Dollinger, P.S. Walsh, and R. Griffith. 1992. Simultaneous amplification and detection of specific DNA sequences. *Biotechnology* 10: 413–417.

Holland, P.M., R.D. Abramson, R. Watson, and D.H. Gelfand. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'→3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proc. Natl. Acad. Sci. USA* 88: 7276–7280.

Huang, S.K., H.Q. Xiao, T.J. Kleine, G. Pantoli, H.G. Marsh, L.M. Lichtenstein, and M.C. Liu. 1995a. IL-13 expression at the sites of allergen challenge in patients with asthma. *J. Immun.* 155: 2688–2694.

Huang, S.K., M. Yi, E. Palmai, and D.G. Marsh. 1995b. A dominant T cell receptor beta-chain in response to a short ragweed allergen. *Amb a 5. J. Immun.* 154: 6157–6162.

Kellogg, D.E., J.J. Sninsky, and S. Kwok. 1990. Quantitation of HIV-1 proviral DNA relative to cellular DNA by the polymerase chain reaction. *Anal. Biochem.* 189: 202–208.

Lee, L.G., C.R. Connell, and W. Bloch. 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Res.* 21: 3761–3766.

Livak, K.J., S.J. Flood, J. Marmaro, W. Giusti, and K. Deetz. 1995a. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods Applic.* 4: 357–362.

Livak, K.J., J. Marmaro, and J.A. Todd. 1995b. Towards fully automated genome-wide polymorphism screening [Letter]. *Nature Genet.* 9: 341–342.

Mulder, J., N. McKinney, C. Christopherson, J. Sninsky, L. Greenfield, and S. Kwok. 1994. Rapid and simple PCR assay for quantitation of human immunodeficiency virus type 1 RNA in plasma: Application to acute retroviral infection. *J. Clin. Microbiol.* 32: 292–300.

Pang, S., Y. Koyanagi, S. Miles, C. Wiley, H.V. Vinters, and I.S. Chen. 1990. High levels of unintegrated HIV-1 DNA in brain tissue of AIDS dementia patients. *Nature* 343: 85–89.

Piatak, M.J., K.C. Luk, B. Williams, and J.D. Lifson. 1993. Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *BioTechniques* 14: 70–81.

Piatak, M.J., M.S. Saag, L.C. Yang, S.J. Clark, J.C. Kappes, K.C. Luk, B.H. Hahn, G.M. Shaw, and J.D. Lifson. 1993b. High levels of HIV-1 in plasma during all stages of infection determined by competitive PCR. *Science* 259: 1749–1754.

Prud'homme, G.J., D.H. Kono, and A.N. Theofilopoulos. 1995. Quantitative polymerase chain reaction analysis reveals marked overexpression of interleukin-1 beta, interleukin-1 and interferon-gamma mRNA in the lymph nodes of lupus-prone mice. *Mol. Immunol.* 32: 495–503.

Ratcliffe, L. 1995. A commentary on the practical applications of competitive PCR. *Genome Res.* 5: 91–94.

Sharp, P.A., A.J. Berk, and S.M. Berget. 1980. Transcription maps of adenovirus. *Methods Enzymol.* 65: 750–768.

Slamon, D.J., G.M. Clark, S.G. Wong, W.J. Levin, A. Ullrich, and W.L. McGuire. 1987. Human breast cancer: Correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235: 177–182.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98: 503–517.

Tan, X., X. Sun, G.R. Gonzalez, and W. Hauck. 1994. PAF and TNF increase the precursor of NF-kappa B p50 mRNA in mouse intestine: Quantitative analysis by competitive PCR. *Biochim. Biophys. Acta* 1226: 157–162.

Thomas, P.S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci.* 77: 5201–5205.

Williams, S., C. Schwer, A. Krishnarao, C. Held, B. Kaiser, and P.M. Williams. 1996. Quantitative competitive PCR: Analysis of amplified products of the HIV-1 *pol* gene by capillary electrophoresis with laser induced fluorescence detection. *Anal. Biochem.* (In press).

Received June 3, 1996; accepted in revised form July 29, 1996.

HellerEhrman
ATTORNEYS

methods. Peptides AENK or AEQK were dissolved in water, made isotonic with NaCl and diluted into RPMI growth medium. T-cell-proliferation assays were done essentially as described[20,21]. Briefly, after antigen pulsing (30 μg ml^{-1} TTCF) with tetrapeptides (1–2 mg ml^{-1}), PBMCs or EBV-B cells were washed in PBS and fixed for 45 s in 0.05% glutaraldehyde. Glycine was added to a final concentration of 0.1M and the cells were washed five times in RPMI 1640 medium containing 1% FCS before co-culture with T-cell clones in round-bottom 96-well microtitre plates. After 48 h, the cultures were pulsed with 1 μCi of ^{3}H-thymidine and harvested for scintillation counting 16 h later. Predigestion of native TTCF was done by incubating 200 μg TTCF with 0.25 μg pig kidney legumain in 500 μl 50 mM citrate buffer, pH 5.5, for 1 h at 37 °C.

Glycopeptide digestions. The peptides HIDNEEDI, HIDN(*N*-glucosamine) EEDI and HIDNESDI, which are based on the TTCF sequence, and QQQHLFGSNVTDCSGNFCLFR(KKK), which is based on human transferrin, were obtained by custom synthesis. The three C-terminal lysine residues were added to the natural sequence to aid solubility. The transferrin glycopeptide QQQHLFGSNVTDCSGNFCLFR was prepared by tryptic (Promega) digestion of 5 mg reduced, carboxy-methylated human transferrin followed by concanavalin A chromatography[41]. Glycopeptides corresponding to residues 622–642 and 421–452 were isolated by reverse-phase HPLC and identified by mass spectrometry and N-terminal sequencing. The lyophilized transferrin-derived peptides were redissolved in 50 mM sodium acetate, pH 5.5, 10 mM dithiothreitol, 20% methanol. Digestions were performed for 3 h at 30 °C with 5–50 m*U* ml^{-1} pig kidney legumain or B-cell AEP. Products were analysed by HPLC or MALDI-TOF mass spectrometry using a matrix of 10 mg ml^{-1} α-cyanocinnamic acid in 50% acetonitrile/0.1% TFA and a PerSeptive Biosystems Elite STR mass spectrometer set to linear or reflector mode. Internal standardization was obtained with a matrix ion of 568.13 mass units.

Received 29 September; accepted 3 November 1998.

1. Chen, J. M. *et al.* Cloning, isolation, and characterization of mammalian legumain, an asparaginyl endopeptidase. *J. Biol. Chem.* 272, 8090–8098 (1997).
2. Kembhavi, A. A., Buttle, D. J., Knight, C. G. & Barrett, A. J. The two cysteine endopeptidases of legume seeds: purification and characterization by use of specific fluorometric assays. *Arch. Biochem. Biophys.* 303, 208–213 (1993).
3. Dalton, J. P., Hola-Jamriska, L. & Bridley, P. J. Asparaginyl endopeptidase activity in adult *Schistosoma mansoni*. *Parasitology* 111, 575–580 (1995).
4. Bennett, K. *et al.* Antigen processing for presentation by class II major histocompatibility complex requires cleavage by cathepsin E. *Eur. J. Immunol.* 22, 1517–1524 (1992).
5. Riese, R. J. *et al.* Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* 4, 357–366 (1996).
6. Rodriguez, G. M. & Diment, S. Role of cathepsin D in antigen presentation of ovalbumin. *J. Immunol.* 149, 2894–2898 (1992).
7. Hewitt, E. W. *et al.* Natural processing sites for human cathepsin E and cathepsin D in tetanus toxin: implications for T cell epitope generation. *J. Immunol.* 159, 4693–4699 (1997).
8. Watts, C. Capture and processing of exogenous antigens for presentation on MHC molecules. *Annu. Rev. Immunol.* 15, 821–850 (1997).
9. Chapman, H. A. Endosomal proteases and MHC class II function. *Curr. Opin. Immunol.* 10, 93–102 (1998).
10. Pieters, J. & Miller, J. Endosomal proteases and antigen processing. *Trends Biochem. Sci.* 22, 377–382 (1997).
11. Liu, J. & van Halbeek, H. Complete ^{1}H and ^{13}C resonance assignments of a 21-amino acid glycopeptide prepared from human serum transferrin. *Carbohydr. Res.* 296, 1–21 (1996).
12. Pearson, G. T. & Lockeley, R. M. The instructive role of innate immunity in the acquired immune response. *Science* 272, 50–54 (1996).
13. Medzhitov, R. & Janeway, C. A. J. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 91, 295–298 (1997).
14. Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393, 705–711 (1998).
15. Mocarelli, P. *et al.* N-glycosylation of HIV gp120 may constrain recognition by T lymphocytes. *J. Immunol.* 147, 3128–3132 (1991).
16. Davidson, H. W., West, M. A. & Watts, C. Endocytosis, intracellular trafficking, and processing of membrane IgG and monovalent antigen/membrane IgG complexes in B lymphocytes. *J. Immunol.* 144, 4101–4109 (1990).
17. Barrett, A. J. & Kirschke, H. Cathepsin B, cathepsin H and cathepsin L. *Methods Enzymol.* 80, 535–559 (1981).
18. Makoff, A. J., Oxbrantine, S. P., Smallwood, A. E. & Fairweather, N. F. Expression of tetanus toxin fragment C in *E. coli*: its purification and potential use as a vaccine. *Biotechnology* 7, 1043–1046 (1989).
19. Lane, D. P. & Harlow, E. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1988).
20. Lanzavecchia, A. Antigen-specific interaction between T and B cells. *Nature* 314, 537–539 (1985).
21. Pond, L. & Watts, C. Characterization of transport of newly assembled, T cell-stimulatory MHC class II-peptide complexes from MHC class II compartments to the cell surface. *J. Immunol.* 159, 543–553 (1997).

Acknowledgements. We thank M. Ferguson for helpful discussions and advice; E. Smythe and L. Grayson for advice and technical assistance; B. Spruce, A. Knight and the BTS (Ninewells Hospital) for help with blood monocyte preparation; and our colleagues for many helpful comments on the manuscript. This work was supported by the Wellcome Trust and by an EMBO Long-term fellowship to B. M.

Correspondence and requests for materials should be addressed to C.W. (e-mail: c.watts@dundee.ac.uk).

Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer

Robert M. Pitti*†, Scot A. Marsters*†, David A. Lawrence*†, Margaret Roy*, Frank C. Kischkel*, Patrick Dowd*, Arthur Huang*, Christopher J. Donahue*, Steven W. Sherwood*, Daryl T. Baldwin*, Paul J. Godowski*, William I. Wood*, Austin L. Gurney*, Kenneth J. Hillan*, Robert L. Cohen*, Audrey D. Goddard*, David Botstein‡ & Avi Ashkenazi*

** Departments of Molecular Oncology, Molecular Biology, and Immunology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA*
‡ Department of Genetics, Stanford University, Stanford, California 94305, USA
† These authors contributed equally to this work

Fas ligand (FasL) is produced by activated T cells and natural killer cells and it induces apoptosis (programmed cell death) in target cells through the death receptor Fas/Apo1/CD95 (ref. 1). One important role of FasL and Fas is to mediate immune-cytotoxic killing of cells that are potentially harmful to the organism, such as virus-infected or tumour cells[2]. Here we report the discovery of a soluble decoy receptor, termed decoy receptor 3 (DcR3), that binds to FasL and inhibits FasL-induced apoptosis. The DcR3 gene was amplified in about half of 35 primary lung and colon tumours studied, and DcR3 messenger RNA was expressed in malignant tissue. Thus, certain tumours may escape FasL-dependent immune-cytotoxic attack by expressing a decoy receptor that blocks FasL.

By searching expressed sequence tag (EST) databases, we identified a set of related ESTs that showed homology to the tumour necrosis factor (TNF) receptor (TNFR) gene superfamily[3]. Using the overlapping sequence, we isolated a previously unknown full-length complementary DNA from human fetal lung. We named the protein encoded by this cDNA decoy receptor 3 (DcR3). The cDNA encodes a 300-amino-acid polypeptide that resembles members of the TNFR family (Fig. 1a): the amino terminus contains a leader sequence, which is followed by four tandem cysteine-rich domains (CRDs). Like one other TNFR homologue, osteoprotegerin (OPG)[3], DcR3 lacks an apparent transmembrane sequence, which indicates that it may be a secreted, rather than a membrane-associated, molecule. We expressed a recombinant, histidine-tagged form of DcR3 in mammalian cells; DcR3 was secreted into the cell culture medium, and migrated on polyacrylamide gels as a protein of relative molecular mass 35,000 (data not shown). DcR3 shares sequence identity in particular with OPG (31%) and TNFR2 (29%), and has relatively less homology with Fas (17%). All of the cysteines in the four CRDs of DcR3 and OPG are conserved; however, the carboxy-terminal portion of DcR3 is 101 residues shorter.

We analysed expression of DcR3 mRNA in human tissues by northern blotting (Fig. 1b). We detected a predominant 1.2-kilobase transcript in fetal lung, brain, and liver, and in adult spleen, colon and lung. In addition, we observed relatively high DcR3 mRNA expression in the human colon carcinoma cell line SW480.

To investigate potential ligand interactions of DcR3, we generated a recombinant, Fc-tagged DcR3 protein. We tested binding of DcR3–Fc to human 293 cells transfected with individual TNF-family ligands, which are expressed as type 2 transmembrane proteins (these transmembrane proteins have their N termini in the cytosol). DcR3–Fc showed a significant increase in binding to cells transfected with FasL[4] (Fig. 2a), but not to cells transfected with TNF[5], Apo2L/TRAIL[6,7], Apo3L/TWEAK[8,9], or OPGL/TRANCE/

letters to nature

RANKL[10-12] (data not shown). DcR3–Fc immunoprecipitated shed FasL from FasL-transfected 293 cells (Fig. 2b) and purified soluble FasL (Fig. 2c), as did the Fc-tagged ectodomain of Fas but not TNFR1. Gel-filtration chromatography showed that DcR3–Fc and soluble FasL formed a stable complex (Fig. 2d). Equilibrium analysis indicated that DcR3–Fc and Fas–Fc bound to soluble FasL with a comparable affinity ($K_d = 0.8 \pm 0.2$ and 1.1 ± 0.1 nM, respectively; Fig. 2e), and that DcR3–Fc could block nearly all of the binding of soluble FasL to Fas–Fc (Fig. 2e, inset). Thus, DcR3 competes with Fas for binding to FasL.

To determine whether binding of DcR3 inhibits FasL activity, we tested the effect of DcR3–Fc on apoptosis induction by soluble FasL in Jurkat T leukaemia cells, which express Fas (Fig. 3a). DcR3–Fc and Fas–Fc blocked soluble-FasL-induced apoptosis in a similar dose-dependent manner, with half-maximal inhibition at $\sim 0.1\ \mu g\ ml^{-1}$. Time-course analysis showed that the inhibition did not merely delay cell death, but rather persisted for at least 24 hours (Fig. 3b). We also tested the effect of DcR3–Fc on activation-induced cell death (AICD) of mature T lymphocytes, a FasL-dependent process[1]. Consistent with previous results[13], activation of interleukin-2-stimulated CD4-positive T cells with anti-CD3 antibody increased the level of apoptosis twofold, and Fas–Fc blocked this effect substantially (Fig. 3c); DcR3–Fc blocked the induction of apoptosis to a similar extent. Thus, DcR3 binding blocks apoptosis induction by FasL.

FasL-induced apoptosis is important in elimination of virus-infected cells and cancer cells by natural killer cells and cytotoxic T lymphocytes; an alternative mechanism involves perforin and granzymes[14-16]. Peripheral blood natural killer cells triggered marked cell death in Jurkat T leukaemia cells (Fig. 3d); DcR3–Fc and Fas–Fc each reduced killing of target cells from ~65% to ~30%, with half-maximal inhibition at $\sim 1\ \mu g\ ml^{-1}$; the residual killing was probably mediated by the perforin/granzyme pathway. Thus, DcR3 binding blocks FasL-dependent natural killer cell activity. Higher DcR3–Fc and Fas–Fc concentrations were required to block natural killer cell activity compared with those required to block soluble FasL activity, which is consistent with the greater potency of membrane-associated FasL compared with soluble FasL[17].

Given the role of immune cytotoxic cells in elimination of tumour cells and the fact that DcR3 can act as an inhibitor of FasL, we proposed that DcR3 expression might contribute to the ability of some tumours to escape immune-cytotoxic attack. As genomic amplification frequently contributes to tumorigenesis, we investigated whether the DcR3 gene is amplified in cancer. We analysed DcR3 gene-copy number by quantitative polymerase chain

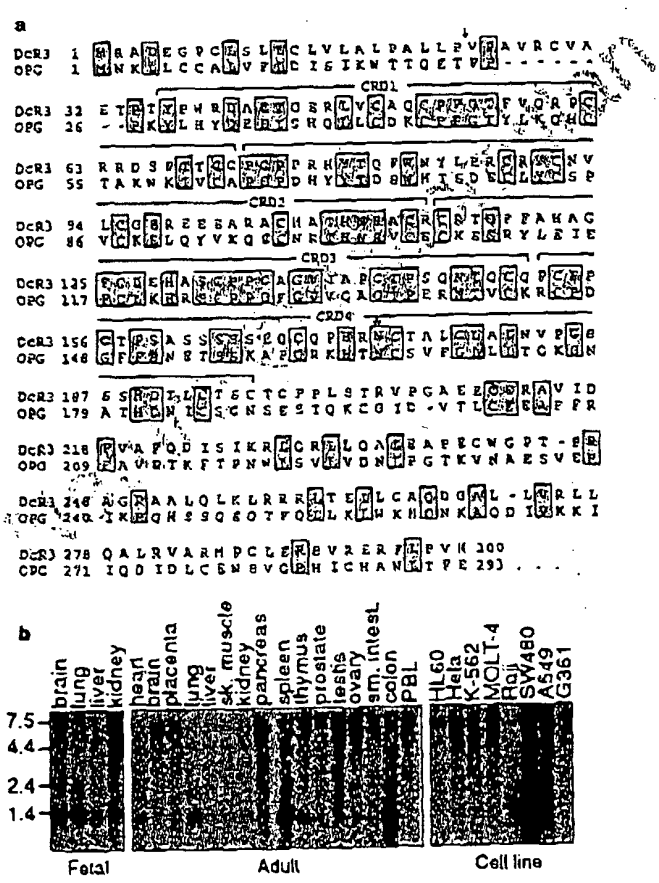


Figure 1 Primary structure and expression of human DcR3. a, Alignment of the amino-acid sequences of DcR3 and of osteoprotegerin (OPG); the C-terminal 101 residues of OPG are not shown. The putative signal cleavage site (arrow), the cysteine-rich domains (CRD 1–4), and the *N*-linked glycosylation site (asterisk) are shown. b, Expression of DcR3 mRNA. Northern hybridization analysis was done using the DcR3 cDNA as a probe and blots of poly(A)$^+$ RNA (Clontech) from human fetal and adult tissues or cancer cell lines. PBL, peripheral blood lymphocyte.

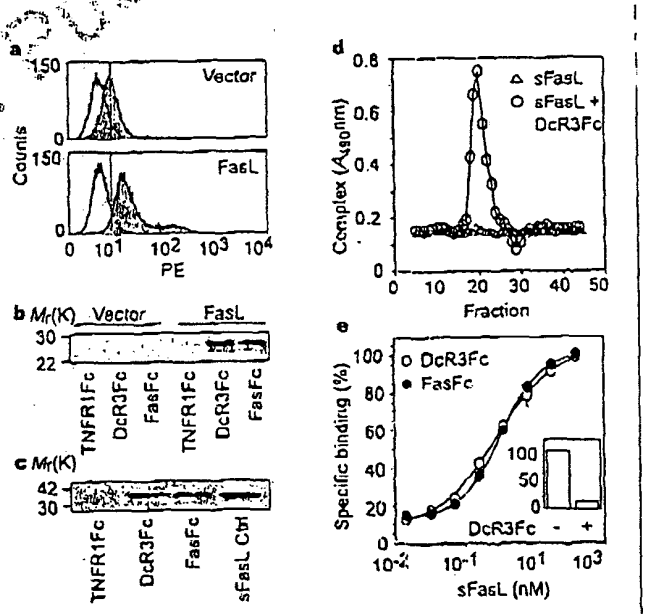


Figure 2 Interaction of DcR3 with FasL. a, 293 cells were transfected with pRK5 vector (top) or with pRK5 encoding full-length FasL (bottom), incubated with DcR3–Fc (solid line, shaded area), TNFR1–Fc (dotted line) or buffer control (dashed line) (the dashed and dotted lines overlap), and analysed for binding by FACS. Statistical analysis showed a significant difference ($P < 0.001$) between the binding of DcR3–Fc to cells transfected with FasL or pRK5. PE, phycoerythrin-labelled cells. b, 293 cells were transfected as in a and metabolically labelled, and cell supernatants were immunoprecipitated with Fc-tagged TNFR1, DcR3 or Fas. c, Purified soluble FasL (sFasL) was immunoprecipitated with TNFR1–Fc, DcR3–Fc or Fas–Fc and visualized by immunoblot with anti-FasL antibody. sFasL was loaded directly for comparison in the right-hand lane. d, Flag-tagged sFasL was incubated with DcR3–Fc or with buffer and resolved by gel filtration; column fractions were analysed in an assay that detects complexes containing DcR3–Fc and sFasL–Flag. e, Equilibrium binding of DcR3–Fc or Fas–Fc to sFasL–Flag. Inset, competition of DcR3–Fc with Fas–Fc for binding to sFasL–Flag.

reaction (PCR)[18] in genomic DNA from 35 primary lung and colon tumours, relative to pooled genomic DNA from peripheral blood leukocytes (PBLs) of 10 healthy donors. Eight of 18 lung tumours and 9 of 17 colon tumours showed DcR3 gene amplification, ranging from 2- to 18-fold (Fig. 4a, b). To confirm this result, we analysed the colon tumour DNAs with three more, independent sets of DcR3-based PCR primers and probes; we observed nearly the same amplification (data not shown).

We then analysed DcR3 mRNA expression in primary tumour tissue sections by *in situ* hybridization. We detected DcR3 expression in 6 out of 15 lung tumours, 2 out of 2 colon tumours, 2 out of 5 breast tumours, and 1 out of 1 gastric tumour (data not shown). A section through a squamous-cell carcinoma of the lung is shown in Fig. 4c. DcR3 mRNA was localized to infiltrating malignant epithelium, but was essentially absent from adjacent stroma, indicating tumour-specific expression. Although the individual tumour specimens that we analysed for mRNA expression and gene amplification were different, the *in situ* hybridization results are consistent with the finding that the DcR3 gene is amplified frequently in tumours. SW480 colon carcinoma cells, which showed abundant DcR3 mRNA expression (Fig. 1b), also had marked DcR3 gene amplification, as shown by quantitative PCR (fourfold) and by Southern blot hybridization (fivefold) (data not shown).

If DcR3 amplification in cancer is functionally relevant, then DcR3 should be amplified more than neighbouring genomic regions that are not important for tumour survival. To test this, we mapped the human DcR3 gene by radiation-hybrid analysis; DcR3 showed linkage to marker AFM216xe7 (T160), which maps to chromosome position 20q13. Next, we isolated from a bacterial artificial chromosome (BAC) library a human genomic clone that carries DcR3, and sequenced the ends of the clone's insert. We then determined, from the nine colon tumours that showed twofold or greater amplification of DcR3, the copy number of the DcR3-flanking sequences (reverse and forward) from the BAC, and of seven genomic markers that span chromosome 20 (Fig. 4d). The DcR3-linked reverse marker showed an average amplification of roughly threefold, slightly less than the approximately fourfold amplification of DcR3; the other markers showed little or no amplification. These data indicate that DcR3 may be at the 'epicentre' of a distal chromosome 20 region that is amplified in colon cancer, consistent with the possibility that DcR3 amplification promotes tumour survival.

Our results show that DcR3 binds specifically to FasL and inhibits FasL activity. We did not detect DcR3 binding to several other TNF-ligand-family members; however, this does not rule out the possibility that DcR3 interacts with other ligands, as do some other TNFR family members, including OPG[2,19].

FasL is important in regulating the immune response; however, little is known about how FasL function is controlled. One mechanism involves the molecule cFLIP, which modulates apoptosis signalling downstream of Fas[20]. A second mechanism involves proteolytic shedding of FasL from the cell surface[17]. DcR3 competes with Fas for

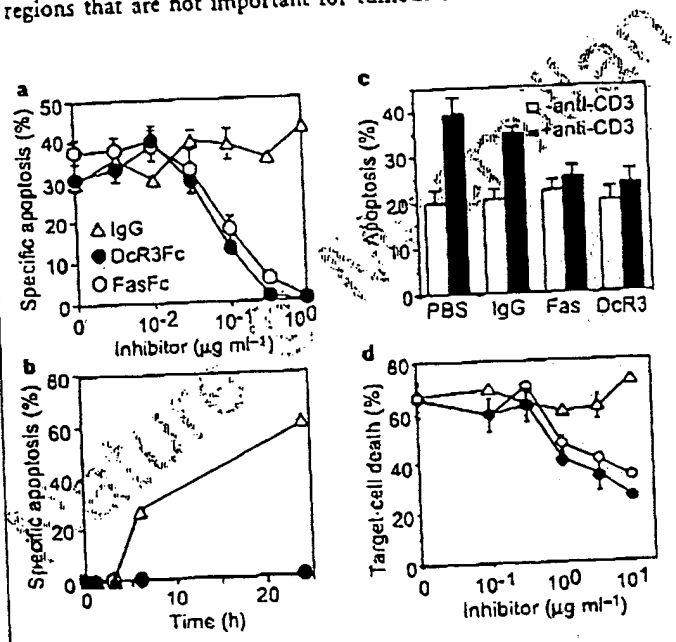


Figure 3 Inhibition of FasL activity by DcR3. a, Human Jurkat T leukaemia cells were incubated with Flag-tagged soluble FasL (sFasL; 5 ng ml^{-1}) oligomerized with anti-Flag antibody (0.1 μg ml^{-1}) in the presence of the proposed inhibitors DcR3-Fc, Fas-Fc or human IgG1 and assayed for apoptosis (mean ± s.e.m. of triplicates). b, Jurkat cells were incubated with sFasL–Flag plus anti-Flag antibody as in a, in presence of 1 μg ml^{-1} DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and apoptosis was determined at the indicated time points. c, Peripheral blood T cells were stimulated with PHA and interleukin-2, followed by control (white bars) or anti-CD3 antibody (filled bars), together with phosphate-buffered saline (PBS), human IgG1, Fas-Fc, or DcR3-Fc (10 μg ml^{-1}). After 16 h, apoptosis of CD4$^+$ cells was determined (mean ± s.e.m. of results from five donors). d, Peripheral blood natural killer cells were incubated with ^{51}Cr-labelled Jurkat cells in the presence of DcR3-Fc (filled circles), Fas-Fc (open circles) or human IgG1 (triangles), and target-cell death was determined by release of ^{51}Cr (mean ± s.d. for two donors, each in triplicate).

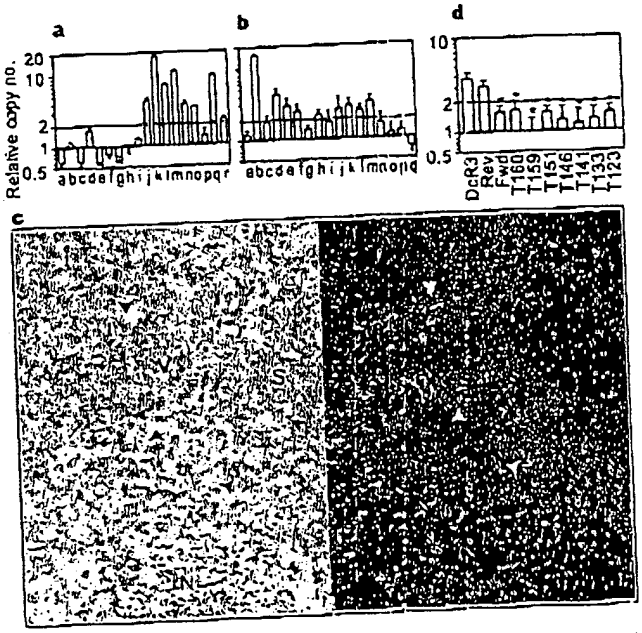


Figure 4 Genomic amplification of DcR3 in tumours. a, Lung cancers, comprising eight adenocarcinomas (c, d, f, g, h, j, k, r), seven squamous-cell carcinomas (a, e, m, n, o, p, q), one non-small-cell carcinoma (b), one small-cell carcinoma (i), and one bronchial adenocarcinoma (l). The data are means ± s.d. of 2 experiments done in duplicate. b, Colon tumours, comprising 17 adenocarcinomas. Data are means ± s.e.m. of five experiments done in duplicate. c, *In situ* hybridization analysis of DcR3 mRNA expression in a squamous-cell carcinoma of the lung. A representative bright-field image (left) and the corresponding dark-field image (right) show DcR3 mRNA over infiltrating malignant epithelium (arrowheads). Adjacent non-malignant stroma (S), blood vessel (V) and necrotic tumour tissue (N) are also shown. d, Average amplification of DcR3 compared with amplification of neighbouring genomic regions (reverse and forward, Rev and Fwd), the DcR3-linked marker T160, and other chromosome-20 markers, in the nine colon tumours showing DcR3 amplification of twofold or more (b). Data are from two experiments done in duplicate. Asterisk indicates $P < 0.01$ for a Student's *t*-test comparing each marker with DcR3.

FasL binding; hence, it may represent a third mechanism of extracellular regulation of FasL activity. A decoy receptor that modulates the function of the cytokine interleukin-1 has been described[21]. In addition, two decoy receptors that belong to the TNFR family, DcR1 and DcR2, regulate the FasL-related apoptosis-inducing molecule Apo2L[22]. Unlike DcR1 and DcR2, which are membrane-associated proteins, DcR3 is directly secreted into the extracellular space. One other secreted TNFR-family member is OPG[3], which shares greater sequence homology with DcR3 (31%) than do DcR1 (17%) or DcR2 (19%); OPG functions as a third decoy for Apo2L[19]. Thus, DcR3 and OPG define a new subset of TNFR-family members that function as secreted decoys to modulate ligands that induce apoptosis. Pox viruses produce soluble TNFR homologues that neutralize specific TNF-family ligands, thereby modulating the antiviral immune response[2]. Our results indicate that a similar mechanism, namely, production of a soluble decoy receptor for FasL, may contribute to immune evasion by certain tumours. □

Methods

Isolation of DcR3 cDNA. Several overlapping ESTs in GenBank (accession numbers AA025672, AA025673 and W67560) and in Lifeseq™ (Incyte Pharmaceuticals: accession numbers 1339238, 1533571, 1533650, 1542861, 1789372 and 2207027) showed similarity to members of the TNFR family. We screened human cDNA libraries by PCR with primers based on the region of EST consensus; fetal lung was positive for a product of the expected size. By hybridization to a PCR-generated probe based on the ESTs, one positive clone (DNA30942) was identified. When searching for potential alternatively spliced forms of DcR3 that might encode a transmembrane protein, we isolated 50 more clones; the coding regions of these clones were identical in size to that of the initial clone (data not shown).

Fc-fusion proteins (immunoadhesins). The entire DcR3 sequence, or the ectodomain of Fas or TNFR1, was fused to the hinge and Fc region of human IgG1, expressed in insect SF9 cells or in human 293 cells, and purified as described[23].

Fluorescence-activated cell sorting (FACS) analysis. We transfected 293 cells using calcium phosphate or Effectene (Qiagen) with pRKS vector or pRKS encoding full-length human FasL (2 μg), together with pRKS encoding CrmA (2 μg) to prevent cell death. After 16 h, the cells were incubated with biotinylated DcR3–Fc or TNFR1–Fc and then with phycoerythrin-conjugated streptavidin (GibcoBRL), and were assayed by FACS. The data were analysed by Kolmogorov–Smirnov statistical analysis. There was some detectable staining of vector-transfected cells by DcR3–Fc; as these cells express little FasL (data not shown), it is possible that DcR3 recognized some other factor that is expressed constitutively on 293 cells.

Immunoprecipitation. Human 293 cells were transfected as above, and metabolically labelled with [^{35}S]cysteine and [^{35}S]methionine (0.5 mCi; Amersham). After 16 h of culture in the presence of z-VAD-fmk (10 μM), the medium was immunoprecipitated with DcR3–Fc, Fas–Fc or TNFR1–Fc (5 μg), followed by protein A–Sepharose (Repligen). The precipitates were resolved by SDS–PAGE and visualized on a phosphorimager (Fuji BAS2000). Alternatively, purified, Flag-tagged soluble FasL (1 μg) (Alexis) was incubated with each Fc-fusion protein (1 μg), precipitated with protein A–Sepharose, resolved by SDS–PAGE and visualized by immunoblotting with rabbit anti-FasL antibody (Oncogene Research).

Analysis of complex formation. Flag-tagged soluble FasL (25 μg) was incubated with buffer or with DcR3–Fc (40 μg) for 1.5 h at 24 °C. The reaction was loaded onto a Superdex 200 HR 10/30 column (Pharmacia) and developed with PBS; 0.6-ml fractions were collected. The presence of DcR3–Fc–FasL complex in each fraction was analysed by placing 100 μl aliquots into microtitre wells precoated with anti-human IgG (Boehringer) to capture DcR3–Fc, followed by detection with biotinylated anti-Flag antibody Bio M2 (Kodak) and streptavidin–horseradish peroxidase (Amersham). Calibration of the column indicated an apparent relative molecular mass of the complex of 420K (data not shown), which is consistent with a stoichiometry of two DcR3–Fc homodimers to two soluble FasL homotrimers.

Equilibrium binding analysis. Microtitre wells were coated with anti-human IgG, blocked with 2% BSA in PBS. DcR3–Fc or Fas–Fc was added, followed by serially diluted Flag-tagged soluble FasL. Bound ligand was detected with anti-Flag antibody as above. In the competition assay, Fas–Fc was immobilized as above, and the wells were blocked with excess IgG1 before addition of Flag-tagged soluble FasL plus DcR3–Fc.

T-cell AICD. CD3$^+$ lymphocytes were isolated from peripheral blood of individual donors using anti-CD3 magnetic beads (Miltenyi Biotech), stimulated with phytohaemagglutinin (PHA; 2 μg ml^{-1}) for 24 h, and cultured in the presence of interleukin-2 (100 U ml^{-1}) for 5 days. The cells were plated in wells coated with anti-CD3 antibody (Pharmingen) and analysed for apoptosis 16 h later by FACS analysis of annexin-V-binding of CD4$^+$ cells[24].

Natural killer cell activity. Natural killer cells were isolated from peripheral blood of individual donors using anti-CD56 magnetic beads (Miltenyi Biotech), and incubated for 16 h with ^{51}Cr-loaded Jurkat cells at an effector-to-target ratio of 1:1 in the presence of DcR3–Fc, Fas–Fc or human IgG1. Target-cell death was determined by release of ^{51}Cr in effector–target co-cultures relative to release of ^{51}Cr by detergent lysis of equal numbers of Jurkat cells.

Gene-amplification analysis. Surgical specimens were provided by J. Kern (lung tumours) and P. Quirke (colon tumours). Genomic DNA was extracted (Qiagen) and the concentration was determined using Hoechst dye 33258 intercalation fluorometry. Amplification was determined by quantitative PCR[18] using a TaqMan instrument (ABI). The method was validated by comparison of PCR and Southern hybridization data for the Myc and HER-2 oncogenes (data not shown). Gene-specific primers and fluorogenic probes were designed on the basis of the sequence of DcR3 or of nearby regions identified on a BAC carrying the human DcR3 gene: alternatively, primers and probes were based on Stanford Human Genome Center marker AFM218xe7 (T160), which is linked to DcR3 (likelihood score = 5.4), SHGC-36268 (T159), the nearest available marker which maps to ~500 kilobases from T160, and five extra markers that span chromosome 20. The DcR3-specific primer sequences were 5′-CTTCTTCGCGCACGCTG-3′ and 5′-ATCACGCCGGCACCAG-3′ and the fluorogenic probe sequence was 5′-(FAM-ACACGATGCGTGCTCCAAGCAG AAp-(TAMARA), where FAM is 5′-fluorescein phosphoramidite. Relative gene-copy numbers were derived using the formula $2^{(\Delta CT)}$, where ΔCT is the difference in amplification cycles required to detect DcR3 in peripheral blood lymphocyte DNA compared to test DNA.

Received 24 September; accepted 6 November 1998.

1. Nagata, S. Apoptosis by death factor. *Cell* 88, 355–365 (1997).
2. Smith, C. A., Farrah, T. & Goodwin, R. G. The TNF receptor superfamily of cellular and viral proteins: activation, costimulation, and death. *Cell* 76, 959–962 (1994).
3. Simonet, W. S. *et al.* Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* 89, 309–319 (1997).
4. Suda, T., Takahashi, T., Golstein, P. & Nagata, S. Molecular cloning and expression of Fas ligand, a novel member of the TNF family. *Cell* 75, 1169–1178 (1993).
5. Pennica, D. *et al.* Human tumour necrosis factor: precursor structure, expression and homology to lymphotoxin. *Nature* 312, 724–729 (1984).
6. Pitti, R. M. *et al.* Induction of apoptosis by Apo-2 ligand, a new member of the tumor necrosis factor receptor family. *J. Biol. Chem.* 271, 12687–12690 (1996).
7. Wiley, S. R. *et al.* Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* 3, 673–682 (1995).
8. Marsters, S. A. *et al.* Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* 8, 525–528 (1998).
9. Chicheportiche, Y. *et al.* TWEAK, a new secreted ligand in the TNF family that weakly induces apoptosis. *J. Biol. Chem.* 272, 32401–32410 (1997).
10. Wong, B. R. *et al.* TRANCE is a novel ligand of the TNFR family that activates c-Jun-N-terminal kinase in T cells. *J. Biol. Chem.* 272, 25190–25194 (1997).
11. Anderson, D. M. *et al.* A homolog of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* 390, 175–179 (1997).
12. Lacey, D. L. *et al.* Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* 93, 165–176 (1998).
13. Ohteln, J., Walczak, H., Baumler, C., Debatin, K. M. & Krammer, P. H. Autocrine T-cell suicide mediated by Apo-1/(Fas/CD95). *Nature* 375, 438–441 (1995).
14. Arase, H., Arase, N. & Saito, T. Fas-mediated cytotoxicity by freshly isolated natural killer cells. *J. Exp. Med.* 181, 1235–1238 (1995).
15. Medvedev, A. E. *et al.* Regulation of Fas and Fas ligand expression in NK cells by cytokines and the involvement of Fas ligand in NK/LAK cell-mediated cytotoxicity. *Cytokine* 9, 394–404 (1997).
16. Moretta, A. Mechanisms in cell-mediated cytotoxicity. *Cell* 90, 13–18 (1997).
17. Tanaka, M., Itai, T., Adachi, M. & Nagata, S. Downregulation of Fas ligand by shedding. *Nature Med.* 4, 31–36 (1998).
18. Gelmini, S. *et al.* Quantitative PCR-based homogeneous assay with fluorogenic probes to measure c-erbB-2 oncogene amplification. *Clin. Chem.* 43, 752–758 (1997).
19. Emery, J. G. *et al.* Osteoprotegerin is a receptor for the cytotoxic ligand TRAIL. *J. Biol. Chem.* 273, 14363–14367 (1998).
20. Walloch, D. Placing death under control. *Nature* 388, 121–125 (1997).
21. Colotta, F. *et al.* Interleukin-1 type II receptor: a decoy target for IL-1 that is regulated by IL-4. *Science* 261, 472–475 (1993).

22. Ashkenazi, A. & Dixit, V. M. Death receptors: signaling and modulation. *Science* 281, 1305–1308 (1998).
23. Ashkenazi, A. & Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* 9, 195–200 (1997).
24. Marsters, S. et al. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* 6, 750–752 (1996).

Acknowledgements. We thank C. Clark, D. Pennica and V. Dixit for comments, and J. Kern and P. Quirke for tumour specimens.

Correspondence and requests for materials should be addressed to A.A. (e-mail: aa@gene.com). The GenBank accession number for the DcR3 cDNA sequence is AF104419.

Crystal structure of the ATP-binding subunit of an ABC transporter

Li-Wei Hung*, Iris Xiaoyan Wang†, Kishiko Nikaido†, Pei-Qi Liu†, Giovanna Ferro-Luzzi Ames† & Sung-Hou Kim*‡

* *E. O. Lawrence Berkeley National Laboratory, † Department of Molecular and Cell Biology, and ‡ Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, USA*

ABC transporters (also known as traffic ATPases) form a large family of proteins responsible for the translocation of a variety of compounds across membranes of both prokaryotes and eukaryotes[1]. The recently completed *Escherichia coli* genome sequence revealed that the largest family of paralogous *E. coli* proteins is composed of ABC transporters[2]. Many eukaryotic proteins of medical significance belong to this family, such as the cystic fibrosis transmembrane conductance regulator (CFTR), the P-glycoprotein (or multidrug-resistance protein) and the heterodimeric transporter associated with antigen processing (Tap1–Tap2). Here we report the crystal structure at 1.5 Å resolution of HisP, the ATP-binding subunit of the histidine permease, which is an ABC transporter from *Salmonella typhimurium*. We correlate the details of this structure with the biochemical, genetic and biophysical properties of the wild-type and several mutant HisP proteins. The structure provides a basis for understanding properties of ABC transporters and of defective CFTR proteins.

ABC transporters contain four structural domains: two nucleotide-binding domains (NBDs), which are highly conserved throughout the family, and two transmembrane domains[1]. In prokaryotes these domains are often separate subunits which are assembled into a membrane-bound complex; in eukaryotes the domains are generally fused into a single polypeptide chain. The periplasmic histidine permease of *S. typhimurium* and *E. coli*[1,3–5] is a well-characterized ABC transporter that is a good model for this superfamily. It consists of a membrane-bound complex, $HisQMP_2$, which comprises integral membrane subunits, HisQ and HisM, and two copies of HisP, the ATP-binding subunit. HisP, which has properties intermediate between those of integral and peripheral membrane proteins[6], is accessible from both sides of the membrane, presumably by its interaction with HisQ and HisM[6]. The two HisP subunits form a dimer, as shown by their cooperativity in ATP hydrolysis[5], the requirement for both subunits to be present for activity[4], and the formation of a HisP dimer upon chemical cross-linking. Soluble HisP also forms a dimer[3]. HisP has been purified and characterized in an active soluble form[3] which can be reconstituted into a fully active membrane-bound complex[4].

The overall shape of the crystal structure of the HisP monomer is that of an 'L' with two thick arms (arm I and arm II); the ATP-binding pocket is near the end of arm I (Fig. 1). A six-stranded β-sheet (β3 and β8–β12) spans both arms of the L, with a domain of a α- plus β-type structure (β1, β2, β4–β7, α1 and α2) on one side (within arm I) and a domain of mostly α-helices (α3–α9) on the

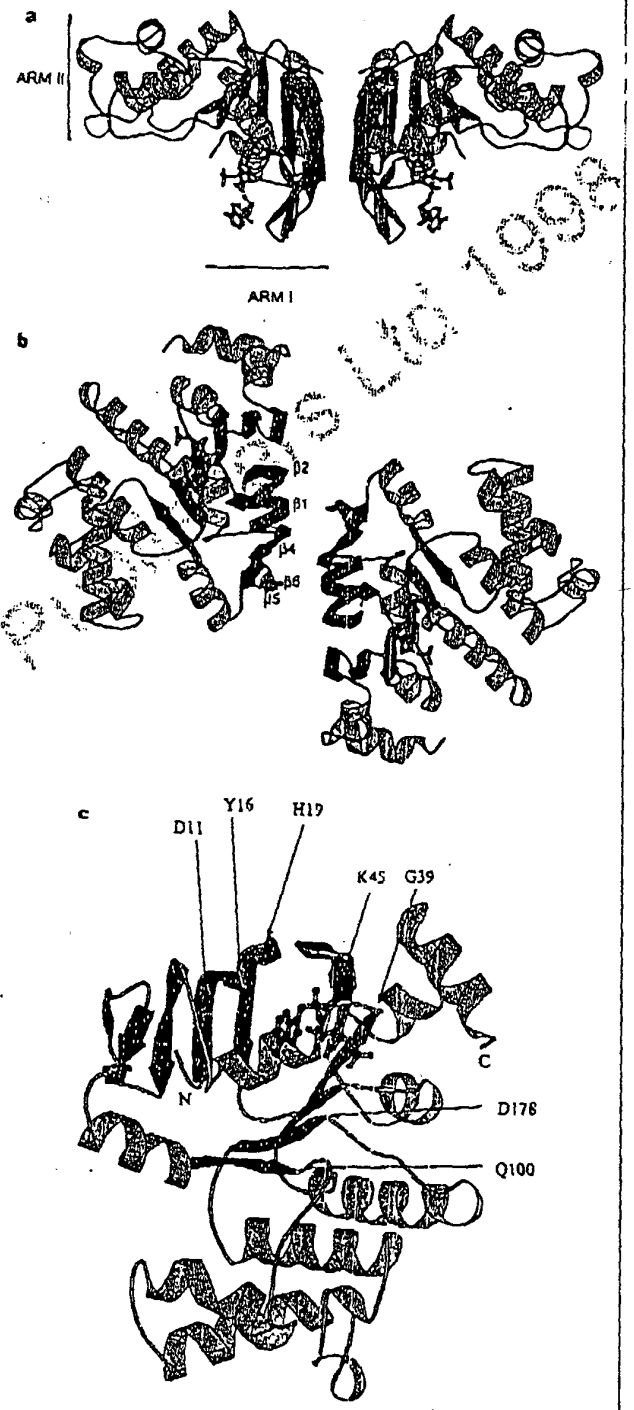


Figure 1 Crystal structure of HisP. **a**, View of the dimer along an axis perpendicular to its two-fold axis. The top and bottom of the dimer are suggested to face towards the periplasmic and cytoplasmic sides, respectively (see text). The thickness of arm II is about 25 Å, comparable to that of membrane. α-Helices are shown in orange and β-sheets in green. **b**, View along the two-fold axis of the HisP dimer, showing the relative displacement of the monomers not apparent in **a**. The β-strands at the dimer interface are labelled. **c**, View of one monomer from the bottom of arm I, as shown in **a**, towards arm II, showing the ATP-binding pocket. **a**–**c**, The protein and the bound ATP are in 'ribbon' and 'ball-and-stick' representations, respectively. Key residues discussed in the text are indicated in **c**. These figures were prepared with MOLSCRIPT[28]. N, amino terminus; C, C terminus.

HellerEhrman
ATTORNEYS

Int. J. Cancer: 78, 661–666 (1998)

Publication of the International Union Against Cancer
Publication de l'Union Internationale Contre le Cancer

NOVEL APPROACH TO QUANTITATIVE POLYMERASE CHAIN REACTION USING REAL-TIME DETECTION: APPLICATION TO THE DETECTION OF GENE AMPLIFICATION IN BREAST CANCER

Ivan BIÈCHE[1,2], Martine OLIVI[1], Marie-Hélène CHAMPÈME[2], Dominique VIDAUD[1], Rosette LIDEREAU[2] and Michel VIDAUD[1]*

[1]*Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, Paris, France*
[2]*Laboratoire d'Oncogénétique, Centre René Huguenin, St-Cloud, France*

Gene amplification is a common event in the progression of human cancers, and amplified oncogenes have been shown to have diagnostic, prognostic and therapeutic relevance. A kinetic quantitative polymerase-chain-reaction (PCR) method, based on fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real-time, was used to quantify gene amplification in tumor DNA. Reactions are characterized by the point during cycling when PCR amplification is still in the exponential phase, rather than the amount of PCR product accumulated after a fixed number of cycles. None of the reaction components is limited during the exponential phase, meaning that values are highly reproducible in reactions starting with the same copy number. This greatly improves the precision of DNA quantification. Moreover, real-time PCR does not require post-PCR sample handling, thereby preventing potential PCR-product carry-over contamination; it possesses a wide dynamic range of quantification and results in much faster and higher sample throughput. The real-time PCR method, was used to develop and validate a simple and rapid assay for the detection and quantification of the 3 most frequently amplified genes (*myc*, *ccnd1* and *erbB2*) in breast tumors. Extra copies of *myc*, *ccnd1* and *erbB2* were observed in 10, 23 and 15%, respectively, of 108 breast-tumor DNA; the largest observed numbers of gene copies were 4.6, 18.6 and 15.1, respectively. These results correlated well with those of Southern blotting. The use of this new semi-automated technique will make molecular analysis of human cancers simpler and more reliable, and should find broad applications in clinical and research settings. *Int. J. Cancer* 78:661–666, 1998.
© 1998 Wiley-Liss, Inc.

Gene amplification plays an important role in the pathogenesis of various solid tumors, including breast cancer, probably because over-expression of the amplified target genes confers a selective advantage. The first technique used to detect genomic amplification was cytogenetic analysis. Amplification of several chromosome regions, visualized either as extrachromosomal double minutes (dmins) or as integrated homogeneously staining regions (HSRs), are among the main visible cytogenetic abnormalities in breast tumors. Other techniques such as comparative genomic hybridization (CGH) (Kallioniemi *et al.*, 1994) have also been used in broad searches for regions of increased DNA copy numbers in tumor cells, and have revealed some 20 amplified chromosome regions in breast tumors. Positional cloning efforts are underway to identify the critical gene(s) in each amplified region. To date, genes known to be amplified frequently in breast cancers include *myc* (8q24), *ccnd1* (11q13), and *erbB2* (17q12-q21) (for review, see Bièche and Lidereau, 1995).

Amplification of the *myc*, *ccnd1*, and *erbB2* proto-oncogenes should have clinical relevance in breast cancer, since independent studies have shown that these alterations can be used to identify sub-populations with a worse prognosis (Berns *et al.*, 1992; Schuuring *et al.*, 1992; Slamon *et al.*, 1987). Muss *et al.* (1994) suggested that these gene alterations may also be useful for the prediction and assessment of the efficacy of adjuvant chemotherapy and hormone therapy.

However, published results diverge both in terms of the frequency of these alterations and their clinical value. For instance, over 500 studies in 10 years have failed to resolve the controversy surrounding the link suggested by Slamon *et al.* (1987) between *erbB2* amplification and disease progression. These discrepancies are partly due to the clinical, histological and ethnic heterogeneity of breast cancer, but technical considerations are also probably involved.

Specific genes (DNA) were initially quantified in tumor cells by means of blotting procedures such as Southern and slot blotting. These batch techniques require large amounts of DNA (5–10 µg/reaction) to yield reliable quantitative results. Furthermore, meticulous care is required at all stages of the procedures to generate blots of sufficient quality for reliable dosage analysis. Recently, PCR has proven to be a powerful tool for quantitative DNA analysis, especially with minimal starting quantities of tumor samples (small, early-stage tumors and formalin-fixed, paraffin-embedded tissues).

Quantitative PCR can be performed by evaluating the amount of product either after a given number of cycles (end-point quantitative PCR) or after a varying number of cycles during the exponential phase (kinetic quantitative PCR). In the first case, an internal standard distinct from the target molecule is required to ascertain PCR efficiency. The method is relatively easy but implies generating, quantifying and storing an internal standard for each gene studied. Nevertheless, it is the most frequently applied method to date.

One of the major advantages of the kinetic method is its rapidity in quantifying a new gene, since no internal standard is required (an external standard curve is sufficient). Moreover, the kinetic method has a wide dynamic range (at least 5 orders of magnitude), giving an accurate value for samples differing in their copy number. Unfortunately, the method is cumbersome and has therefore been rarely used. It involves aliquot sampling of each assay mix at regular intervals and quantifying, for each aliquot, the amplification product. Interest in the kinetic method has been stimulated by a novel approach using fluorescent TaqMan methodology and a new instrument (ABI Prism 7700 Sequence Detection System) capable of measuring fluorescence in real time (Gibson *et al.*, 1996; Heid *et al.*, 1996). The TaqMan reaction is based on the 5′ nuclease assay first described by Holland *et al.* (1991). The latter uses the 5′ nuclease activity of Taq polymerase to cleave a specific fluorogenic oligonucleotide probe during the extension phase of PCR. The approach uses dual-labeled fluorogenic hybridization probes (Lee *et al.*, 1993). One fluorescent dye, co-valently linked to the 5′ end of the oligonucleotide, serves as a reporter [FAM (*i.e.*, 6-carboxy-fluorescein)] and its emission spectrum is quenched by a second fluorescent dye, TAMRA (*i.e.*, 6-carboxy-tetramethyl-rhodamine) attached to the 3′ end. During the extension phase of the PCR

Grant sponsors: Association Pour la Recherche sur le Cancer and Ministère de l'Enseignement Supérieur et de la Recherche.

*Correspondence to: Laboratoire de Génétique Moléculaire, Faculté des Sciences Pharmaceutiques et Biologiques de Paris, 4 Avenue de l'Observatoire, F-75006 Paris, France. Fax: (33)1-1407-1754.
E-mail: mvidaud@tesser.fr

Received 2 May 1998; Revised 30 June 1998

cycle, the fluorescent hybridization probe is hydrolyzed by the 5'-3' nucleolytic activity of DNA polymerase. Nuclease degradation of the probe releases the quenching of FAM fluorescence emission, resulting in an increase in peak fluorescence emission. The fluorescence signal is normalized by dividing the emission intensity of the reporter dye (FAM) by the emission intensity of a reference dye (*i.e.*, ROX, 6-carboxy-X-rhodamine) included in TaqMan buffer, to obtain a ratio defined as the Rn (normalized reporter) for a given reaction tube. The use of a sequence detector enables the fluorescence spectra of all 96 wells of the thermal cycler to be measured continuously during PCR amplification.

The real-time PCR method offers several advantages over other current quantitative PCR methods (Celi *et al.*, 1994): *(i)* the probe-based homogeneous assay provides a real-time method for detecting only specific amplification products, since specific hybridation of both the primers and the probe is necessary to generate a signal; *(ii)* the C_t (threshold cycle) value used for quantification is measured when PCR amplification is still in the log phase of PCR product accumulation. This is the main reason why C_t is a more reliable measure of the starting copy number than are end-point measurements, in which a slight difference in a limiting component can have a drastic effect on the amount of product; *(iii)* use of C_t values gives a wider dynamic range (at least 5 orders of magnitude), reducing the need for serial dilution; *(iv)* The real-time PCR method is run in a closed-tube system and requires no post-PCR sample handling, thus avoiding potential contamination; *(v)* the system is highly automated, since the instrument continuously measures fluorescence in all 96 wells of the thermal cycler during PCR amplification and the corresponding software processes, and analyzes the fluorescence data; *(vi)* the assay is rapid, as results are available just one minute after thermal cycling is complete; *(vii)* the sample throughput of the method is high, since 96 reactions can be analyzed in 2 hr.

Here, we applied this semi-automated procedure to determine the copy numbers of the 3 most frequently amplified genes in breast tumors (*myc*, *ccnd1* and *erb*B2), as well as 2 genes (*alb* and *app*) located in a chromosomic region in which no genetic changes have been observed in breast tumors. The results for 108 breast tumors were compared with previous Southern-blot data for the same samples.

MATERIAL AND METHODS

Tumor and blood samples

Samples were obtained from 108 primary breast tumors removed surgically from patients at the Centre René Huguenin; none of the patients had undergone radiotherapy or chemotherapy. Immediately after surgery, the tumor samples were placed in liquid nitrogen until extraction of high-molecular-weight DNA. Patients were included in this study if the tumor sample used for DNA preparation contained more than 60% of tumor cells (histological analysis). A blood sample was also taken from 18 of the same patients.

DNA was extracted from tumor tissue and blood leukocytes according to standard methods.

Real-time PCR

Theoretical basis. Reactions are characterized by the point during cycling when amplification of the PCR product is first detected, rather than by the amount of PCR product accumulated after a fixed number of cycles. The higher the starting copy number of the genomic DNA target, the earlier a significant increase in fluorescence is observed. The parameter C_t (threshold cycle) is defined as the fractional cycle number at which the fluorescence generated by cleavage of the probe passes a fixed threshold above baseline. The target gene copy number in unknown samples is quantified by measuring C_t and by using a standard curve to determine the starting copy number. The precise amount of genomic DNA (based on optical density) and its quality (*i.e.*, lack of extensive degradation) are both difficult to assess. We therefore also quantified a control gene *(alb)* mapping to chromosome region 4q11-q13, in which no genetic alterations have been found in breast tumor DNA by means of CGH (Kallioniemi *et al.*, 1994).

Thus, the ratio of the copy number of the target gene to the copy number of the *alb* gene normalizes the amount and quality of genomic DNA. The ratio defining the level of amplification is termed "N", and is determined as follows:

$$N = \frac{\text{copy number of target gene (}\textit{app, myc, ccnd1, erb}\text{B2)}}{\text{copy number of reference gene (}\textit{alb}\text{)}}.$$

Primers, probes, reference human genomic DNA and PCR consumables. Primers and probes were chosen with the assistance of the computer programs Oligo 4.0 (National Biosciences, Plymouth, MN), EuGene (Daniben Systems, Cincinnati, OH) and Primer Express (Perkin-Elmer Applied Biosystems, Foster City, CA).

Primers were purchased from DNAgency (Malvern, PA) and probes from Perkin-Elmer Applied Biosystems.

Nucleotide sequences for the oligonucleotide hybridization probes and primers are available on request.

The TaqMan PCR Core reagent kit, MicroAmp optical tubes, and MicroAmp caps were from Perkin-Elmer Applied Biosystems.

Standard-curve construction. The kinetic method requires a standard curve. The latter was constructed with serial dilutions of specific PCR products, according to Piatak *et al.* (1993). In practice, each specific PCR product was obtained by amplifying 20 ng of a standard human genomic DNA (Boehringer, Mannheim, Germany) with the same primer pairs as those used later for real-time quantitative PCR. The 5 PCR products were purified using MicroSpin S-400 HR columns (Pharmacia, Uppsala, Sweden) electrophorezed through an acrylamide gel and stained with ethidium bromide to check their quality. The PCR products were then quantified spectrophotometrically and pooled, and serially diluted 10-fold in mouse genomic DNA (Clontech, Palo Alto, CA) at a constant concentration of 2 ng/μl. The standard curve used for real-time quantitative PCR was based on serial dilutions of the pool of PCR products ranging from 10^{-7} (10^5 copies of each gene) to 10^{-10} (10^2 copies). This series of diluted PCR products was aliquoted and stored at −80°C until use.

The standard curve was validated by analyzing 2 known quantities of calibrator human genomic DNA (20 ng and 50 ng).

PCR amplification. Amplification mixes (50 μl) contained the sample DNA (around 20 ng, around 6600 copies of disomic genes), 10× TaqMan buffer (5 μl), 200 μM dATP, dCTP, dGTP, and 400 μM dUTP, 5 mM $MgCl_2$, 1.25 units of AmpliTaq Gold, 0.5 units of AmpErase uracil N-glycosylase (UNG), 200 nM each primer and 100 nM probe. The thermal cycling conditions comprised 2 min at 50°C and 10 min at 95°C. Thermal cycling consisted of 40 cycles at 95°C for 15 s and 65°C for 1 min. Each assay included: a standard curve (from 10^5 to 10^2 copies) in duplicate, a no-template control, 20 ng and 50 ng of calibrator human genomic DNA (Boehringer) in triplicate, and about 20 ng of unknown genomic DNA in triplicate (26 samples can thus be analyzed on a 96-well microplate). All samples with a coefficient of variation (CV) higher than 10% were retested.

All reactions were performed in the ABI Prism 7700 Sequence Detection System (Perkin-Elmer Applied Biosystems), which detects the signal from the fluorogenic probe during PCR.

Equipment for real-time detection. The 7700 system has a built-in thermal cycler and a laser directed via fiber optical cables to each of the 96 sample wells. A charge-coupled-device (CDD) camera collects the emission from each sample and the data are analyzed automatically. The software accompanying the 7700 system calculates C_t and determines the starting copy number in the samples.

Determination of gene amplification. Gene amplification was calculated as described above. Only samples with an N value higher than 2 were considered to be amplified.

RESULTS

To validate the method, real-time PCR was performed on genomic DNA extracted from 108 primary breast tumors, and 18 normal leukocyte DNA samples from some of the same patients. The target genes were the *myc*, *ccnd1* and *erbB2* proto-oncogenes, and the β-amyloid precursor protein gene (*app*), which maps to a chromosome region (21q21.2) in which no genetic alterations have been found in breast tumors (Kallioniemi *et al.*, 1994). The reference disomic gene was the albumin gene (*alb*, chromosome 4q11-q13).

Validation of the standard curve and dynamic range of real-time PCR

The standard curve was constructed from PCR products serially diluted in genomic mouse DNA at a constant concentration of 2 ng/μl. It should be noted that the 5 primer pairs chosen to analyze the 5 target genes do not amplify genomic mouse DNA (data not shown). Figure 1 shows the real-time PCR standard curve for the *alb* gene. The dynamic range was wide (at least 4 orders of magnitude), with samples containing as few as 10^2 copies or as many as 10^5 copies.

Copy-number ratio of the 2 reference genes (app and alb)

The *app* to *alb* copy-number ratio was determined in 18 normal leukocyte DNA samples and all 108 primary breast-tumor DNA

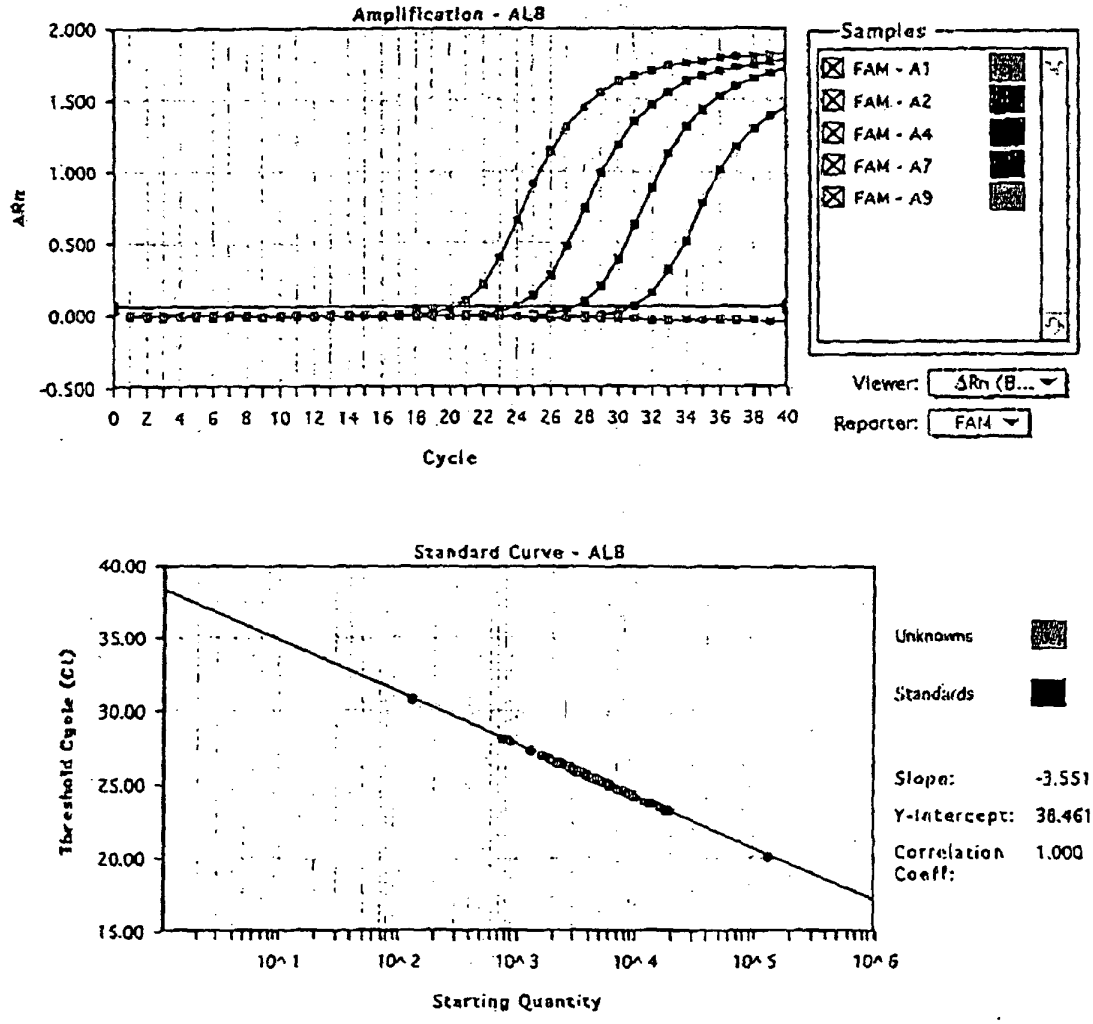


FIGURE 1 – Albumin (*alb*) gene dosage by real-time PCR. Top: Amplification plots for reactions with starting *alb* gene copy number ranging from 10^5 (A9), 10^4 (A7), 10^3 (A4) to 10^2 (A2) and a no-template control (A1). Cycle number is plotted *vs.* change in normalized reporter signal (ΔRn). For each reaction tube, the fluorescence signal of the reporter dye (FAM) is divided by the fluorescence signal of the passive reference dye (ROX), to obtain a ratio defined as the normalized reporter signal (Rn). ΔRn represents the normalized reporter signal (Rn) minus the baseline signal established in the first 15 PCR cycles. ΔRn increases during PCR as *alb* PCR product copy number increases until the reaction reaches a plateau. C_t (threshold cycle) represents the fractional cycle number at which a significant increase in Rn above a baseline signal (horizontal black line) can first be detected. Two replicate plots were performed for each standard sample, but the data for only one are shown here. Bottom: Standard curve plotting log starting copy number *vs.* C_t (threshold cycle). The black dots represent the data for standard samples plotted in duplicate and the red dots the data for unknown genomic DNA samples plotted in triplicate. The standard curve shows 4 orders of linear dynamic range.

samples. We selected these 2 genes because they are located in 2 chromosome regions (*app*, 21q21.2; *alb*, 4q11-q13) in which no obvious genetic changes (including gains or losses) have been observed in breast cancers (Kallioniemi *et al.*, 1994). The ratio for the 18 normal leukocyte DNA samples fell between 0.7 and 1.3 (mean 1.02 ± 0.21), and was similar for the 108 primary breast-tumor DNA samples (0.6 to 1.6, mean 1.06 ± 0.25), confirming that *alb* and *app* are appropriate reference disomic genes for breast-tumor DNA. The low range of the ratios also confirmed that the nucleotide sequences chosen for the primers and probes were not polymorphic, as mismatches of their primers or probes with the subject's DNA would have resulted in differential amplification.

myc, ccnd1 *and* erbB2 *gene dose in normal leukocyte DNA*

To determine the cut-off point for gene amplification in breast-cancer tissue, 18 normal leukocyte DNA samples were tested for the gene dose (N), calculated as described in "Material and Methods". The N value of these samples ranged from 0.5 to 1.3 (mean 0.84 ± 0.22) for *myc*, 0.7 to 1.6 (mean 1.06 ± 0.23) for *ccnd1* and 0.6 to 1.3 (mean 0.91 ± 0.19) for *erb*B2. Since N values for *myc*, *ccnd1* and *erb*B2 in normal leukocyte DNA consistently fell between 0.5 and 1.6, values of 2 or more were considered to represent gene amplification in tumor DNA.

myc, ccnd1 *and* erbB2 *gene dose in breast-tumor DNA*

myc, *ccnd1* and *erb*B2 gene copy numbers in the 108 primary breast tumors are reported in Table I. Extra copies of *ccnd1* were more frequent (23%, 25/108) than extra copies of *erb*B2 (15%, 16/108) and *myc* (10%, 11/108), and ranged from 2 to 18.6 for *ccnd1*, 2 to 15.1 for *erb*B2, and only 2 to 4.6 for the *myc* gene. Figure 2 and Table II represent tumors in which the *ccnd1* gene was amplified 16-fold (T145), 6-fold (T133) and non-amplified (T118). The 3 genes were never found to be co-amplified in the same tumor. *erb*B2 and *ccnd1* were co-amplified in only 3 cases, *myc* and *ccnd1* in 2 cases and *myc* and *erb*B2 in 1 case. This favors the hypothesis that gene amplifications are independent events in breast cancer. Interestingly, 5 tumors showed a decrease of at least 50% in the *erb*B2 copy number (N < 0.5), suggesting that they bore deletions of the 17q21 region (the site of *erb*B2). No such decrease in copy number was observed with the other 2 proto-oncogenes.

Comparison of gene dose determined by real-time quantitative PCR and Southern-blot analysis

Southern-blot analysis of *myc*, *ccnd1* and *erb*B2 amplifications had previously been done on the same 108 primary breast tumors. A perfect correlation between the results of real-time PCR and Southern blot was obtained for tumors with high copy numbers (N ≥ 5). However, there were cases (1 *myc*, 6 *ccnd1* and 4 *erb*B2) in which real-time PCR showed gene amplification whereas Southern-blot did not, but these were mainly cases with low extra copy numbers (N from 2 to 2.9).

DISCUSSION

The clinical applications of gene amplification assays are currently limited, but would certainly increase if a simple, standardized and rapid method were perfected. Gene amplification status has been studied mainly by means of Southern blotting, but this method is not sensitive enough to detect low-level gene amplification nor accurate enough to quantify the full range of amplification values. Southern blotting is also time-consuming, uses radioactive reagents and requires relatively large amounts of high-quality genomic DNA, which means it cannot be used routinely in many laboratories. An amplification step is therefore required to determine the copy number of a given target gene from minimal quantities of tumor DNA (small early-stage tumors, cytopuncture specimens or formalin-fixed, paraffin-embedded tissues).

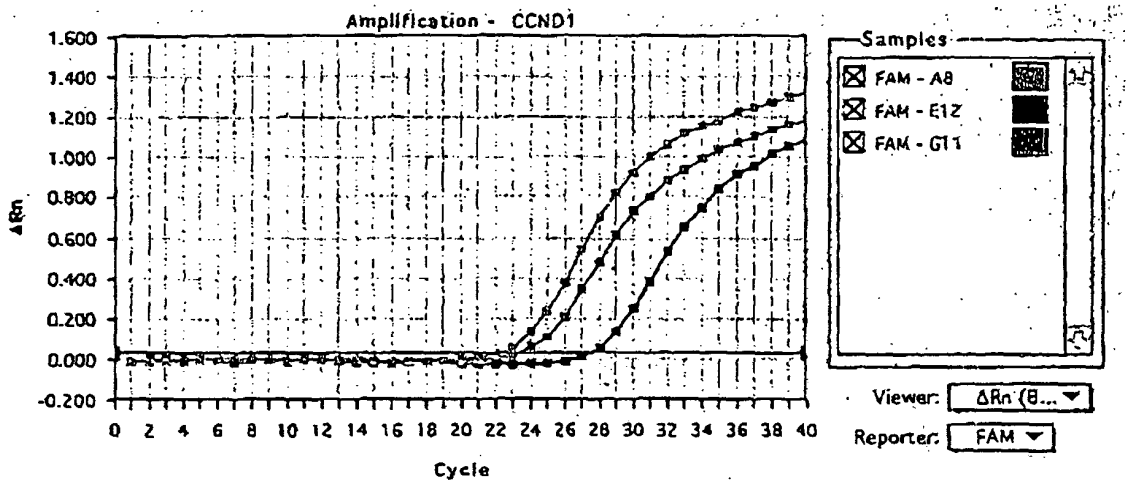
In this study, we validated a PCR method developed for the quantification of gene over-representation in tumors. The method, based on real-time analysis of PCR amplification, has several advantages over other PCR-based quantitative assays such as competitive quantitative PCR (Celi *et al.*, 1994). First, the real-time PCR method is performed in a closed-tube system, avoiding the risk of contamination by amplified products. Re-amplification of carryover PCR products in subsequent experiments can also be prevented by using the enzyme uracil N-glycosylase (UNG) (Longo *et al.*, 1990). The second advantage is the simplicity and rapidity of sample analysis, since no post-PCR manipulations are required. Our results show that the automated method is reliable. We found it possible to determine, in triplicate, the number of copies of a target gene in more than 100 tumors per day. Third, the system has a linear dynamic range of at least 4 orders of magnitude, meaning that samples do not have to contain equal starting amounts of DNA. This technique should therefore be suitable for analyzing formalin-fixed, paraffin-embedded tissues. Fourth, and above all, real-time PCR makes DNA quantification much more precise and reproducible, since it is based on C_t values rather than end-point measurement of the amount of accumulated PCR product. Indeed, the ABI Prism 7700 Sequence Detection System enables C_t to be calculated when PCR amplification is still in the exponential phase and when none of the reaction components is rate-limiting. The within-run CV of the C_t value for calibrator human DNA (5 replicates) was always below 5%, and the between-assay precision in 5 different runs was always below 10% (data not shown). In addition, the use of a standard curve is not absolutely necessary, since the copy number can be determined simply by comparing the C_t ratio of the target gene with that of reference genes. The results obtained by the 2 methods (with and without a standard curve) are similar in our experiments (data not shown). Moreover, unlike competitive quantitative PCR, real-time PCR does not require an internal control (the design and storage of internal controls and the validation of their amplification efficiency is laborious).

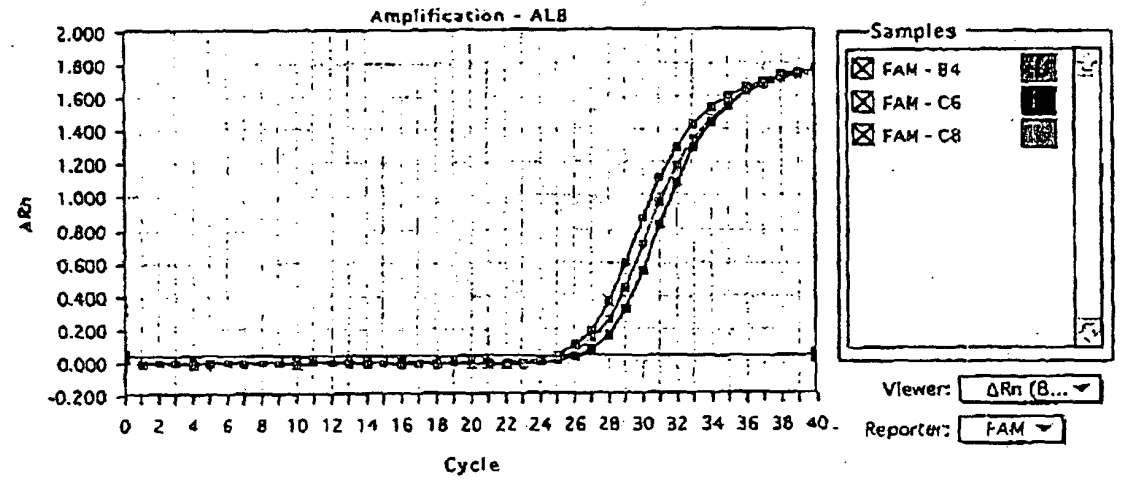
The only potential disavantage of real-time PCR, like all other PCR-based methods and solid-matrix blotting techniques (Southern blots and dot blots) is that is cannot avoid dilution artifacts inherent in the extraction of DNA from tumor cells contained in heterogeneous tissue specimens. Only FISH and immunohistochemistry can measure alterations on a cell-by-cell basis (Pauletti *et al.*, 1996; Slamon *et al.*, 1989). However, FISH requires expensive equipment and trained personnel and is also time-consuming. Moreover, FISH does not assess gene expression and therefore cannot detect cases in which the gene product is over-expressed in the absence of gene amplification, which will be possible in the future by real-time quantitative RT-PCR. Immunohistochemistry is subject to considerable variations in the hands of different teams, owing to alterations of target proteins during the procedure, the different primary antibodies and fixation methods used and the criteria used to define positive staining.

The results of this study are in agreement with those reported in the literature. (*i*) Chromosome regions 4q11-q13 and 21q21.2 (which bear *alb* and *app*, respectively) showed no genetic alterations in the breast-cancer samples studied here, in keeping with the results of CGH (Kallioniemi *et al.*, 1994). (*ii*) We found that amplifications of these 3 oncogenes were independent events, as reported by other teams (Berns *et al.*, 1992; Borg *et al.*, 1992). (*iii*) The frequency and degree of *myc* amplification in our breast tumor DNA series were lower than those of *ccnd1* and *erb*B2 amplification, confirming the findings of Borg *et al.* (1992) and Courjal *et al.* (1997). (*iv*) The maxima of *ccnd1* and *erb*B2 over-representation were 18-fold and 15-fold, also in keeping with earlier results (about

TABLE I – DISTRIBUTION OF AMPLIFICATION LEVEL (N) FOR *myc*, *ccnd1* AND *erb*B2 GENES IN 108 HUMAN BREAST TUMORS

Gene	Amplification level (N)			
	<0.5	0.5–1.9	2–4.9	≥5
myc	0	97 (89.8%)	11 (10.2%)	0
ccnd1	0	83 (76.9%)	17 (15.7%)	8 (7.4%)
*erb*B2	5 (4.6%)	87 (80.6%)	8 (7.4%)	8 (7.4%)





Tumor	CCND1 C_t	CCND1 Copy number	ALB C_t	ALB Copy number
T118	27.3	4605	26.5	4365
T133	23.2	61659	25.2	10092
T145	22.1	125892	25.6	7762

FIGURE 2 – *ccnd1* and alb gene dosage by real-time PCR in 3 breast tumor samples: T118 (E12, C6, black squares), T133 (G11, B4, red squares) and T145 (A8, C8, blue squares). Given the C_t of each sample, the initial copy number is inferred from the standard curve obtained during the same experiment. Triplicate plots were performed for each tumor sample, but the data for only one are shown here. The results are shown in Table II.

30-fold maximum) (Berns *et al.*, 1992; Borg *et al.*, 1992; Courjal *et al.*, 1997). (v) The *erb*B2 copy numbers obtained with real-time PCR were in good agreement with data obtained with other quantitative PCR-based assays in terms of the frequency and degree of amplification (An *et al.*, 1995; Deng *et al.*, 1996; Valeron *et al.*, 1996). Our results also correlate well with those recently published by Gelmini *et al.* (1997), who used the TaqMan system to measure *erb*B2 amplification in a small series of breast tumors ($n = 25$), but with an instrument (LS-50B luminescence spectrometer, Perkin-Elmer Applied Biosystems) which only allows end-

TABLE II – EXAMPLES OF *ccnd1* GENE DOSAGE RESULTS FROM 3 BREAST TUMORS[1]

Tumor	*ccnd1*			*alb*			Nccnd1/alb
	Copy number	Mean	SD	Copy number	Mean	SD	
T118	4525			4223			
	4605	4603	77	4365	4325	89	1.06
	4678			4387			
T133	59821			9787			
	61659	61100	1111	10092	10137	375	6.03
	61821			10533			
T145	128563			7321			
	125892	125392	3448	7762	7672	316	16.34
	121722			7933			

[1]For each sample, 3 replicate experiments were performed and the mean and the standard deviation (SD) was determined. The level of *ccnd1* gene amplification (*Nccnd1/alb*) is determined by dividing the average *ccnd1* copy number value by the average *alb* copy number value.

point measurement of fluorescence intensity. Here we report *myc* and *ccnd1* gene dosage in breast cancer by means of quantitative PCR. (vi) We found a high degree of concordance between real-time quantitative PCR and Southern blot analysis in terms of gene amplification, especially for samples with high copy numbers (≥5-fold). The slightly higher frequency of gene amplification (especially *ccnd1* and *erb*B2) observed by means of real-time quantitative PCR as compared with Southern-blot analysis may be explained by the higher sensitivity of the former method. However, we cannot rule out the possibility that some tumors with a few extra gene copies observed in real-time PCR had additional copies of an arm or a whole chromosome (trisomy, tetrasomy or polysomy) rather than true gene amplification. These 2 types of genetic alteration (polysomy and gene amplification) could be easily distinguished in the future by using an additional probe located on the same chromosome arm, but some distance from the target gene. It is noteworthy that high gene copy numbers have the greatest prognostic significance in breast carcinoma (Borg *et al.*, 1992; Slamon *et al.*, 1987).

Finally, this technique can be applied to the detection of gene deletion as well as gene amplification. Indeed, we found a decreased copy number of *erb*B2 (but not of the other 2 proto-oncogenes) in several tumors; *erb*B2 is located in a chromosome region (17q21) reported to contain both deletions and amplifications in breast cancer (Bièche and Lidereau, 1995).

In conclusion, gene amplification in various cancers can be used as a marker of pre-neoplasia, also for early diagnosis of cancer, staging, prognostication and choice of treatment. Southern blotting is not sufficiently sensitive, and FISH is lengthy and complex. Real-time quantitative PCR overcomes both these limitations, and is a sensitive and accurate method of analyzing large numbers of samples in a short time. It should find a place in routine clinical gene dosage.

ACKNOWLEDGEMENTS

RL is a research director at the Institut National de la Santé et de la Recherche Médicale (INSERM). We thank the staff of the Centre René Huguenin for assistance in specimen collection and patient care.

REFERENCES

AN, H.X., NIEDERACHER, D., BECKMANN, M.W., GÖHRING, U.J., SCHARL, A., PICARD, F., VAN ROEYEN, C., SCHNÜRCH, H.G. and BENDER, H.G., *erb*B2 gene amplification detected by fluorescent differential polymerase chain reaction in paraffin-embedded breast carcinoma tissues. *Int. J. Cancer (Pred. Oncol.)*, 64, 291–297 (1995).

BERNS, E.M.J.J., KLIJN, J.G.M., VAN PUTTEN, W.L.J., VAN STAVEREN, I.L., PORTENGEN, H. and FOEKENS, J.A., c-*myc* amplification is a better prognostic factor than HER2/*neu* amplification in primary breast cancer. *Cancer Res.*, 52, 1107–1113 (1992).

BIÈCHE, I. and LIDEREAU, R., Genetic alterations in breast cancer. *Genes Chrom. Cancer*, 14, 227–251 (1995).

BORG, A., BALDETORP, B., FERNO, M., OLSSON, H. and SIGURDSSON, H., c-*myc* amplification is an independent prognostic factor in post-menopausal breast cancer. *Int. J. Cancer*, 51, 687–691 (1992).

CELI, F.S., COHEN, M.M., ANTONARAKIS, S.E., WERTHEIMER, E., ROTH, J. and SHULDINER, A.R., Determination of gene dosage by a quantitative adaptation of the polymerase chain reaction (gd-PCR): rapid detection of deletions and duplications of gene sequences. *Genomics*, 21, 304–310 (1994).

COURJAL, F., CUNY, M., SIMONY-LAFONTAINE, J., LOUASSON, G., SPEISER, P., ZEILLINGER, R., RODRIGUEZ, C. and THEILLET, C., Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: definition of phenotypic groups. *Cancer Res.*, 57, 4360–4367 (1997).

DENG, G., YU, M., CHEN, L.C., MOORE, D., KURISU, W., KALLIONIEMI, A., WALDMAN, F.M., COLLINS, C. and SMITH, H.S., Amplifications of oncogene *erb*B-2 and chromosome 20q in breast cancer determined by differentially competitive polymerase chain reaction. *Breast Cancer Res. Treat.*, 40, 271–281 (1996).

GELMINI, S., ORLANDO, C., SESTINI, R., VONA, G., PINZANI, P., RUOCCO, L. and PAZZAGLI, M., Quantitative polymerase chain reaction-based homogeneous assay with fluorogenic probes to measure c-*erb*B-2 oncogene amplification. *Clin. Chem.*, 43, 752–758 (1997).

GIBSON, U.E.M., HEID, C.A. and WILLIAMS, P.M., A novel method for real-time quantitative RT-PCR. *Genome Res.*, 6, 995–1001 (1996).

HEID, C.A., STEVENS, J., LIVAK, K.J. and WILLIAMS, P.M., Real-time quantitative PCR. *Genome Res.*, 6, 986–994 (1996).

HOLLAND, P.M., ABRAMSON, R.D., WATSON, R. and GELFAND, D.H., Detection of specific polymerase chain reaction product by utilizing the 5' to 3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proc. nat. Acad. Sci. (Wash.)*, 88, 7276–7280 (1991).

KALLIONIEMI, A., KALLIONIEMI, O.P., PIPER, J., TANNER, M., STOKKES, T., CHEN, L., SMITH, H.S., PINKEL, D., GRAY, J.W. and WALDMAN, F.M., Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. nat. Acad. Sci. (Wash.)*, 91, 2156–2160 (1994).

LEE, L.G., CONNELL, C.R. and BIOCH, W., Allelic discrimination by nick-translation PCR with fluorogenic probe. *Nucleic Acids Res.*, 21, 3761–3766 (1993).

LONGO, N., BERNINGER, N.S. and HARTLEY, J.L., Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene*, 93, 125–128 (1990).

MUSS, H.B., THOR, A.D., BERRY, D.A., KUTE, T., LIU, E.T., KOERNER, P., CIRRINCIONE, C.T., BUDMAN, D.R., WOOD, W.C., BARCOS, M. and HENDERSON, I.C., c-*erb*B-2 expression and response to adjuvant therapy in women with node-positive early breast cancer. *New Engl. J. Med.*, 330, 1260–1266 (1994).

PAULETTI, G., GODOLPHIN, W., PRESS, M.F. and SALMON, D.J., Detection and quantification of HER-2/*neu* gene amplification in human breast cancer archival material using fluorescence *in situ* hybridization. *Oncogene*, 13, 63–72 (1996).

PIATAK, M., LUK, K.C., WILLIAMS, B. and LIFSON, J.D., Quantitative competitive polymerase chain reaction for accurate quantitation of HIV DNA and RNA species. *Biotechniques*, 14, 70–80 (1993).

SCHUURING, E., VERHOEVEN, E., VAN TINTEREN, H., PETERSE, J.L., NUNNIK, B., THUNNISSEN, F.B.J.M., DEVILEE, P., CORNELISSE, C.J., VAN DE VIJVER, M.J., MOOI, W.J. and MICHALIDES, R.J.A.M., Amplification of genes within the chromosome 11q13 region is indicative of poor prognosis in patients with operable breast cancer. *Cancer Res.*, 52, 5229–5234 (1992).

SLAMON, D.J., CLARK, G.M., WONG, S.G., LEVIN, W.S., ULLRICH, A. and MCGUIRE, W.L., Human breast cancer: correlation of relapse and survival with amplification of the HER-2/*neu* oncogene. *Science*, 235, 177–182 (1987).

SLAMON, D.J., GODOLPHIN, W., JONES, L.A., HOLT, J.A., WONG, S.G., KEITH, D.E., LEVIN, W.J., STUART, S.G., UDOVE, J., ULLRICH, A. and PRESS, M.F., Studies of the HER-2/*neu* proto-oncogene in human breast and ovarian cancer. *Science*, 244, 707–712 (1989).

VALERON, P.F., CHIRINO, R., FERNANDEZ, L., TORRES, S., NAVARRO, D., AGUIAR, J., CABRERA, J.J., DIAZ-CHICO, B.N. and DIAZ-CHICO, J.C., Validation of a differential PCR and an ELISA procedure in studying HER-2/*neu* status in breast cancer. *Int. J. Cancer*, 65, 129–133 (1996).

APPENDIX B

```
<first sequence: p1.DNA44804 (length = 598)
<second sequence: p1.holtzman (length = 673)

<597 matches in an overlap of 598: 99.83 percent similarity
<gaps in first sequence: 1 (75 residues), gaps in second sequence: 0
<score: 2895 (Dayhoff PAM 250 matrix, gap penalty = 8 + 4 per residue)
<endgaps not penalized

                10        20        30        40        50        60
p1.DNA44804     MCSRVPLLLPLLLLLALGPGVQGCPSGCQCSQPQTVFCTARQGTTVPRDVPPDTVGLYVF
                ************************************************************
p1.holtzman     MCSRVPLLLPLLLLLALGPGVQGCPSGCQCSQPQTVFCTARQGTTVPRDVPPDTVGLYVF
                10        20        30        40        50        60

                70        80        90
p1.DNA44804     ENGITMLDASSFAGLPGLQLLDLSQNQIAS------------------------------
                *********.**********************
p1.holtzman     ENGITMLDAGSFAGLPGLQLLDLSQNQIASLPSGVFQPLANLSNLDLTANRLHEITNETF
                70        80        90       100       110       120

                                                                  100
p1.DNA44804     ------------------------------------------------LRLPRLLLLDLSHNS
                                                                  ***************
p1.holtzman     RGLRRLERLYLGKNRIRHIQPGAFDTLDRLLELKLQDNELRALPPLRLPRLLLLDLSHNS
                      130       140       150       160       170       180

                  110       120       130       140       150       160
p1.DNA44804     LLALEPGILDTANVEALRLAGLGLQQLDEGLFSRLRNLHDLDVSDNQLERVPPVIRGLRG
                ************************************************************
p1.holtzman     LLALEPGILDTANVEALRLAGLGLQQLDEGLFSRLRNLHDLDVSDNQLERVPPVIRGLRG
                      190       200       210       220       230       240

                  170       180       190       200       210       220
p1.DNA44804     LTRLRLAGNTRIAQLRPEDLAGLAALQELDVSNLSLQALPGDLSGLFPRLRLLAAARNPF
                ************************************************************
p1.holtzman     LTRLRLAGNTRIAQLRPEDLAGLAALQELDVSNLSLQALPGDLSGLFPRLRLLAAARNPF
                      250       260       270       280       290       300

                  230       240       250       260       270       280
p1.DNA44804     NCVCPLSWFGPWVRESHVTLASPEETRCHFPPKNAGRLLLELDYADFGCPATTTTATVPT
                ************************************************************
p1.holtzman     NCVCPLSWFGPWVRESHVTLASPEETRCHFPPKNAGRLLLELDYADFGCPATTTTATVPT
                      310       320       330       340       350       360

                  290       300       310       320       330       340
p1.DNA44804     TRPVVREPTALSSSLAPTWLSPTAPATEAPSPPSTAPPTVGPVPQPQDCPPSTCLNGGTC
                ************************************************************
p1.holtzman     TRPVVREPTALSSSLAPTWLSPTAPATEAPSPPSTAPPTVGPVPQPQDCPPSTCLNGGTC
                      370       380       390       400       410       420

                  350       360       370       380       390       400
p1.DNA44804     HLGTRHHLACLCPEGFTGLYCESQMGQGTRPSPTPVTPRPPRSLTLGIEPVSPTSLRVGL
                ************************************************************
p1.holtzman     HLGTRHHLACLCPEGFTGLYCESQMGQGTRPSPTPVTPRPPRSLTLGIEPVSPTSLRVGL
                      430       440       450       460       470       480

                  410       420       430       440       450       460
p1.DNA44804     QRYLQGSSVQLRSLRLTYRNLSGPDKRLVTLRLPASLAEYTVTQLRPNATYSVCVMPLGP
                ************************************************************
p1.holtzman     QRYLQGSSVQLRSLRLTYRNLSGPDKRLVTLRLPASLAEYTVTQLRPNATYSVCVMPLGP
                      490       500       510       520       530       540

                  470       480       490       500       510       520
```

```
p1.DNA44804     GRVPEGE[redacted]GEAHTPPAVHSNHAPVTQAREGNLP[redacted]PALAAVLLAALAAVGAAYCVR
                ******* [redacted] ********************************.  **********************
p1.holtzman     GRVPEGEEACGEAHTPPAVHSNHAPVTQAREGNLPLLIAPALAAVLLAALAAVGAAYCVR
                       550       560       570       580       590       600

                  530       540       550       560       570       580
p1.DNA44804     RGRAMAAAAQDKGQVGPGAGPLELEGVKVPLEPGPKATEGGGEALPSGSECEVPLMGFPG
                ************************************************************
p1.holtzman     RGRAMAAAAQDKGQVGPGAGPLELEGVKVPLEPGPKATEGGGEALPSGSECEVPLMGFPG
                       610       620       630       640       650       660

                  590
p1.DNA44804     PGLQSPLHAKPYI
                *************
p1.holtzman     PGLQSPLHAKPYI
                       670
```

Sequence file: /home/[redacted]va/Molbio/carpenda/temp. ie/p1.holtzman
motifs in /usr/local/seq/libdata/motif.pro

Motif name: N-glycosylation site.
Accession: PS00001;
Motif: N[!P][ST][!P]

101 NLSN
117 NETF
273 NLSL
500 NLSG
528 NATY

Sequence file: /home/[redacted]va/Molbio/carpenda/temp. ie/p1.DNA44804
motifs in /usr/local/seq/libdata/motif.pro

Motif name: N-glycosylation site.
Accession: PS00001;
Motif: N[!P][ST][!P]

198 NLSL
425 NLSG
453 NATY

HMM file: /usr/seqdb/pfam/Pfam_ls
Sequence file: p1.DNA44804

Query: DNA44804 [598 aa]

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
LRR	Leucine Rich Repeat	59.2	8.8e-14	7
LRRCT	Leucine rich repeat C-terminal domain	47.1	4e-10	1
EGF	EGF-like domain	30.0	5.4e-05	1
LRRNT	Leucine rich repeat N-terminal domain	29.8	6.5e-05	1
fn3	Fibronectin type III domain	13.0	0.15	1

Parsed for domains:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
LRRNT	1/1	23	51	..	1	31	[]	29.8	6.5e-05
LRR	1/7	53	76	..	1	25	[]	5.7	2.1e+02
LRR	2/7	77	102	..	1	25	[]	9.4	65
LRR	3/7	118	141	..	1	25	[]	10.4	44
LRR	4/7	142	164	..	1	25	[]	19.1	0.1
LRR	5/7	165	189	..	1	25	[]	11.1	26
LRR	6/7	190	212	..	1	25	[]	12.3	12
LRRCT	1/1	223	275	..	1	54	[]	47.1	4e-10
EGF	1/1	334	366	..	1	45	[]	30.0	5.4e-05
LRR	7/7	415	437	..	1	25	[]	3.1	4.8e+02
fn3	1/1	383	474	..	1	84	[]	13.0	0.15

HMM file: /usr/seqdb/pfam/Pfam_ls
Sequence file: p1.holtzman

Query: holtzman [673 aa]

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
LRR	Leucine Rich Repeat	108.8	1e-28	11
LRRCT	Leucine rich repeat C-terminal domain	47.1	4e-10	1
EGF	EGF-like domain	30.0	5.4e-05	1
LRRNT	Leucine rich repeat N-terminal domain	29.8	6.5e-05	1
fn3	Fibronectin type III domain	13.0	0.15	1

Parsed for domains:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
LRRNT	1/1	23	51	..	1	31	[]	29.8	6.5e-05
LRR	1/11	53	76	..	1	25	[]	6.1	1.9e+02
LRR	2/11	77	100	..	1	25	[]	21.6	0.019
LRR	3/11	101	124	..	1	25	[]	15.6	1.2
LRR	4/11	125	148	..	1	25	[]	18.1	0.21
LRR	5/11	149	169	..	1	25	[]	9.7	58
LRR	6/11	170	192	..	1	25	[]	6.1	1.8e+02
LRR	7/11	193	216	..	1	25	[]	10.4	44
LRR	8/11	217	239	..	1	25	[]	19.1	0.1
LRR	9/11	240	264	..	1	25	[]	11.1	26
LRR	10/11	265	287	..	1	25	[]	12.3	12
LRRCT	1/1	298	350	..	1	54	[]	47.1	4e-10
EGF	1/1	409	441	..	1	45	[]	30.0	5.4e-05
LRR	11/11	490	512	..	1	25	[]	3.1	4.8e+02
fn3	1/1	458	549	..	1	84	[]	13.0	0.15

APPENDIX B

American Journal of Pathology, Vol. 155, No. 3, September 1999

Id-1 and Id-2 Are Overexpressed in Pancreatic Cancer and in Dysplastic Lesions in Chronic Pancreatitis

Haruhisa Maruyama,* Jörg Kleeff,* Stefan Wildi,* Helmut Friess,† Markus W. Büchler,† Mark A. Israel,‡ and Murray Korc*

From the Division of Endocrinology, Diabetes, and Metabolism, Departments of Medicine, Biological Chemistry and Pharmacology, University of California, Irvine, California; the Department of Visceral and Transplantation Surgery,† University of Bern, Bern, Switzerland; and the Preuss Laboratory,‡ Department of Neurological Surgery, University of California, San Francisco, California*

Id proteins antagonize basic helix-loop-helix proteins, inhibit differentiation, and enhance cell proliferation. In this study we compared the expression of Id-1, Id-2, and Id-3 in the normal pancreas, in pancreatic cancer, and in chronic pancreatitis (CP). Northern blot analysis demonstrated that all three Id mRNA species were expressed at high levels in pancreatic cancer samples by comparison with normal or CP samples. Pancreatic cancer cell lines frequently coexpressed all three Ids, exhibiting a good correlation between Id mRNA and protein levels, as determined by immunoblotting with highly specific anti-Id antibodies. Immunohistochemistry using these antibodies demonstrated the presence of faint Id-1 and Id-2 immunostaining in pancreatic ductal cells in the normal pancreas, whereas Id-3 immunoreactivity ranged from weak to strong. In the cancer tissues, many of the cancer cells exhibited abundant Id-1, Id-2, and Id-3 immunoreactivity. Scoring on the basis of percentage of positive cells and intensity of immunostaining indicated that Id-1 and Id-2 were increased significantly in the cancer cells by comparison with the respective controls. Mild to moderate Id immunoreactivity was also seen in the ductal cells in the CP-like areas adjacent to these cells and in the ductal cells of small and interlobular ducts in CP. In contrast, in dysplastic and atypical papillary ducts in CP, Id-1 and Id-2 immunoreactivity was as significantly elevated as in the cancer cells. These findings suggest that increased Id expression may be associated with enhanced proliferative potential of pancreatic cancer cells and of proliferating or dysplastic ductal cells in CP. ***(Am J Pathol 1999, 155:815–822)***

Basic helix-loop-helix (bHLH) proteins play an important role as transcription factors in cellular development, proliferation, and differentiation.[1,2] The basic domain of the bHLHs is required for binding to an E-box DNA sequence, thus promoting transcription of specific target genes. The HLH domain promotes dimer formation with various members of the bHLH protein family.[1,2] Homodimers of the class B family of bHLH proteins, including *Myo*D, *Neuro*D, and numerous other proteins, are known to activate tissue-specific genes.[3–5] These tissue-specific bHLHs typically form heterodimers with widely expressed class A bHLHs, which include proteins encoded by E2A, E2–2, HEB, and other genes (also termed E-proteins).[6–9] These heterodimers activate transcription of genes that are associated with differentiation.

Id genes encode a family of four HLH proteins that lack the basic DNA binding domain.[1,10] They act as dominant-negative HLH proteins by forming high affinity heterodimers with other bHLH proteins, thereby preventing them from binding to DNA and inhibiting transcription of differentiation-associated genes.[10–12] Id gene expression is down-regulated on differentiation in many cell types *in vitro* and *in vivo*.[13–18] In addition, Id proteins seem to be required for cell cycle progression through G_1/S phase in certain cell types, and interaction between Id-2 and pRB is associated with enhanced proliferation in some cell lines *in vitro*.[19–23]

Pancreatic cancer is the fifth leading cause of cancer death in the United States, with a mortality rate that virtually equals its incidence rate.[24] This malignancy is often associated with the overexpression of a variety of mitogenic growth factors and their receptors, and by oncogenic mutations of K-*ras* and inactivation of the p53 tumor suppressor gene.[25] We have recently reported that pancreatic cancers overexpress the HLH protein Id-2, and that enhanced expression of this protein is evident in the cytoplasm of the cancer cells within the pancreatic tumor mass.[26] It is not known, however, whether the expression of other Id proteins is altered in this malignancy, or whether their expression is altered in chronic pancreatitis

Contract grant sponsor: National Cancer Institute. Contract grant number: U. S. Public Health Service grant CA-40162.

Accepted for publication May 24, 1999.

Address reprint requests to Dr. Murray Korc, Division of Endocrinology, Diabetes and Metabolism, Medical Sciences I, C240, University of California, Irvine, CA 92697. E-mail: mkorc@uci.edu.

(CP), an inflammatory disease that is characterized by dysplastic ducts, foci of proliferating ductal cells, acinar cell degeneration, and fibrosis.[27] We now report that there is a five- to sixfold increase in Id-1 and Id-2 mRNA levels and a twofold increase in Id-3 mRNA levels in pancreatic cancer by comparison with the normal pancreas. In contrast, overall Id mRNA levels are not increased in CP.

Patients and Methods

Normal human pancreatic tissue samples from 7 male and 5 female donors (median age 41.8 years, range 14–68 years), CP tissues from 13 males and 1 female (median age 42.1 years; range 30–56 years), and pancreatic cancer tissues from 10 male and 6 female donors (median age 62.6 years; range 53–83 years) were obtained through an organ donor program and from surgical specimens from patients with severe symptomatic chronic pancreatitis or pancreatic cancer. A partial duodenopancreatectomy (Whipple/pylorus-preserving Whipple; $n = 13$), a left resection of the pancreas ($n = 2$), or a total pancreatectomy ($n = 1$) were carried out in the pancreatic cancer patients. According to the TNM classification of the Union Internationale Contre le Cancer (UICC) 6 tumors were stage 1, 1 was stage 2, and 9 were stage 3 ductal cell adenocarcinoma. Freshly removed tissue samples were fixed in 10% formaldehyde solution for 12 to 24 hours and paraffin-embedded for histological analysis. In addition, tissue samples were frozen in liquid nitrogen immediately on surgical removal and maintained in −80°C until use for RNA extraction. All studies were approved by the Ethics Committee of the University of Bern, Bern, Switzerland, and by the Human Subjects Committee at the University of California, Irvine, California.

Northern Blot Analysis

Northern blot analysis was carried out as described previously.[26,28] Briefly, total RNA was extracted by the single step acid guanidinium thiocyanate phenol chloroform method. RNA was size-fractionated on 1.2% agarose/1.8 mol/L formaldehyde gels, electrotransferred onto nylon membranes, and cross-linked by UV irradiation. Blots were prehybridized and hybridized with cDNA probes and washed under high stringency conditions. The following cDNA probes were used: a 979-bp human Id-1 cDNA probe, a 440-bp human Id-2 cDNA probe, and a 450-bp human Id-3 cDNA probe, covering the entire coding regions of Id-1, Id-2, and Id-3, respectively. A *Bam*HI 190-bp fragment of mouse 7S cDNA that hybridizes with human cytoplasmic RNA was used to confirm equal RNA loading and transfer. Blots were then exposed at −80°C to Kodak BioMax-MS films and the resulting autoradiographs were scanned to quantify the intensity of the radiographic bands.[26,28] For each sample the ratio of Id mRNA expression to 7S expression was calculated. To compare the relative increase in expression of the respective Id mRNA species in the cancer and CP samples, the same normal samples were used for normal/cancer and normal/CP membranes. The median score for Id-1, Id-2, and Id-3 mRNA levels in these normal samples was set to 100. Statistical analysis was performed with SigmaStat software (Jandel Scientific, San Raphael, CA). The rank sum test was used, and $P < 0.05$ was taken as the level of significance.

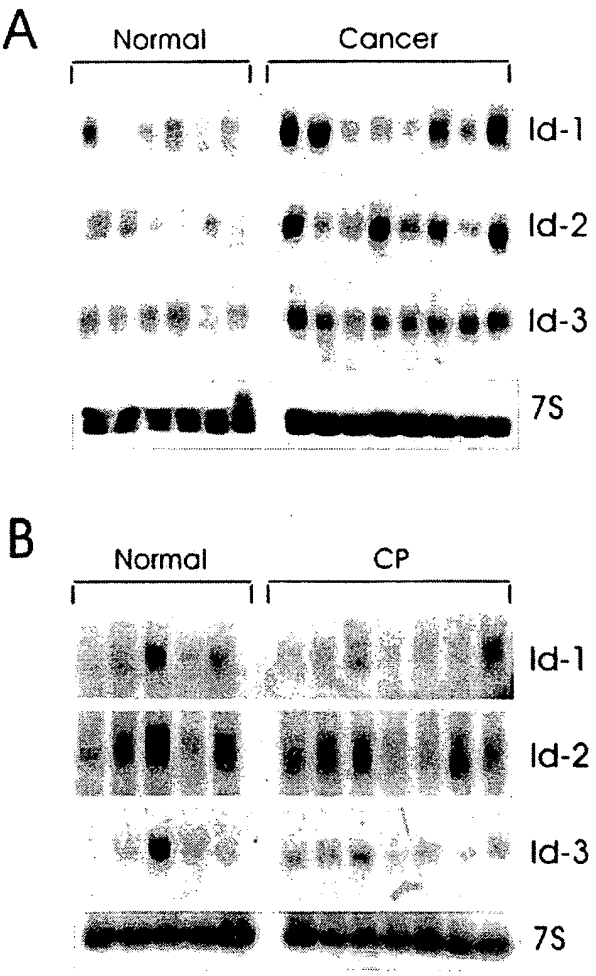


Figure 1. mRNA expression of Id-1, Id-2, and Id-3 in pancreatic cancer and chronic pancreatitis. Total RNA (20 μg/lane) from six normal, eight cancerous, and seven chronic pancreatitis tissue samples were subjected to Northern blot analysis using ^{32}P-labeled cDNA probes (500,000 cpm/ml) specific for Id-1, Id-2, and Id-3, respectively. A 7S cDNA probe (50,000 cpm/ml) was used as a loading and transfer control. Exposure times of the normal/cancer blots were 1 day for all Id probes, and 2 days for the normal/CP blots. Exposure time was 4 hours for mouse 7S cDNA. By comparison with the normal samples, Id-1 and Id-3 mRNA levels were elevated in 8 and 9 cancer samples, respectively, whereas Id-2 was elevated in 6 cancer samples.

Cell Culture and Western Blot Analysis

PANC-1, MIA-PaCa-2, ASPC-1, and CAPAN-1 human pancreatic cell lines were obtained from ATCC (Manassas, VA). COLO-357 human pancreatic cells were a gift from Dr. R. S. Metzger (Durham, NC). Cells were routinely grown in DMEM (COLO-357, MIA-PaCa-2, PANC-1) or RPMI (ASPC-1, CAPAN-1) supplemented with 10% fetal bovine serum, 100 U/ml penicillin, and 100 μg/ml streptomycin. For immunoblot analysis, exponentially growing

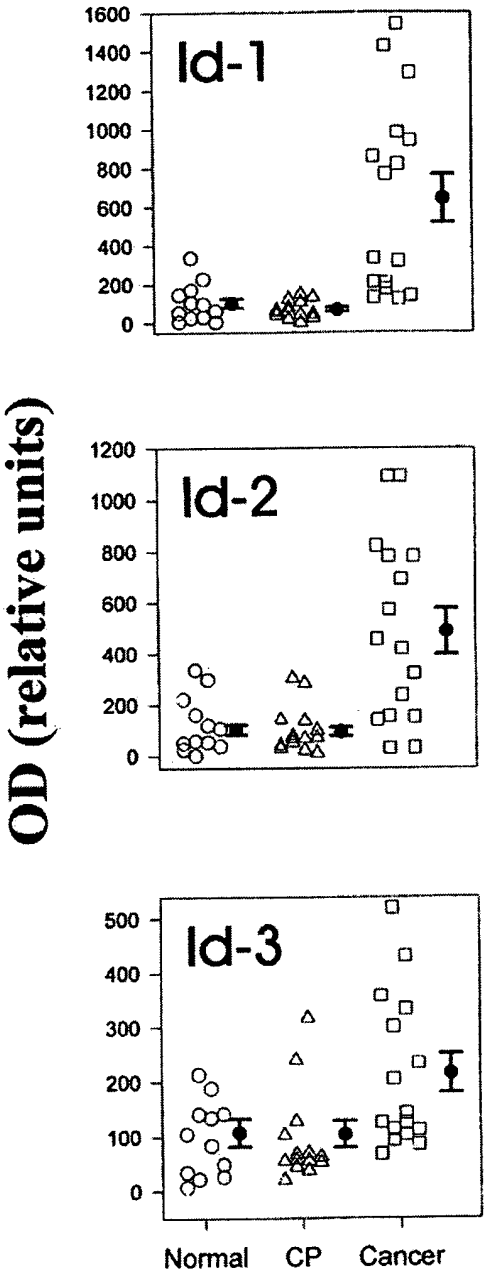


Figure 2. Densitometric analysis of Northern blots. Autoradiographs of Northern blots from 12 normal, 14 CP, and 16 pancreatic cancers were analyzed by densitometry. mRNA levels were determined by calculating the ratio of the optical density for the respective Id mRNA species in relation to the optical density of mouse 7S cDNA. To compare the relative increase in expression of the respective Id mRNA species in the cancer and CP samples, the same normal samples were used for normal/cancer and normal/CP membranes. Normal pancreatic tissues are indicated by circles, CP tissues by triangles, and cancer tissues by squares. Data are expressed as median scores ± SD. By comparison with the normal samples, only the cancer samples exhibited significant increases: 6.5-fold ($P < 0.01$) for Id-1, fivefold ($P < 0.01$) for Id-2, and twofold ($P = 0.027$) for Id-3.

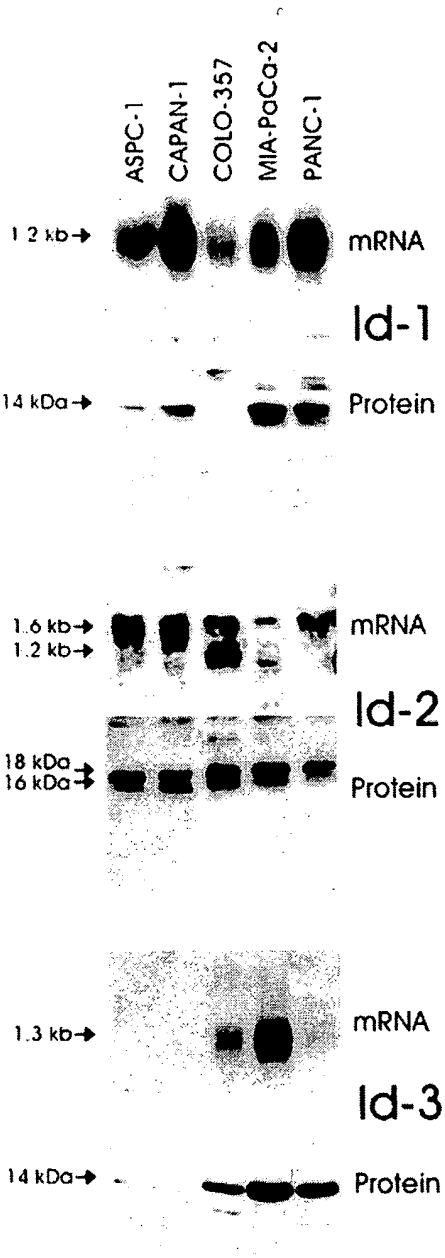


Figure 3. Id mRNA and protein expression in pancreatic cancer cell lines. **Upper panels:** Total RNA (20 μg/lane) from 5 pancreatic cancer cell lines were subjected to Northern blot analysis using ^{32}P-labeled cDNA probes (500,000 cpm/ml) specific for Id-1, Id-2, and Id-3, respectively. Exposure times were 1 day for all Id probes. **Lower panels:** Immunoblotting. Cell lysates (30 μg/lane) were subjected to SDS-PAGE. Membranes were probed with specific Id-1, Id-2, and Id-3 antibodies. Visualization was performed by enhanced chemiluminescence.

cells (60–70% confluent) were solubilized in lysis buffer containing 50 mmol/L Tris-HCl, pH 7.4, 150 mmol/L NaCl, 1 mmol/L EDTA, 1 μg/ml pepstatin A, 1 mmol/L phenylmethylsulfonyl fluoride (PMSF), and 1% Triton X-100. Proteins were subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), transferred to Immobilon P membranes, and incubated for 90 minutes with the indicated antibodies and for 60 minutes with secondary antibodies against rabbit IgG. Visualization was performed by enhanced chemiluminescence.

Immunohistochemistry

Specific rabbit anti-human Id-1 (C-20), Id-2 (C-20), and Id-3 (C-20; all from Santa Cruz Biotechnology, Santa

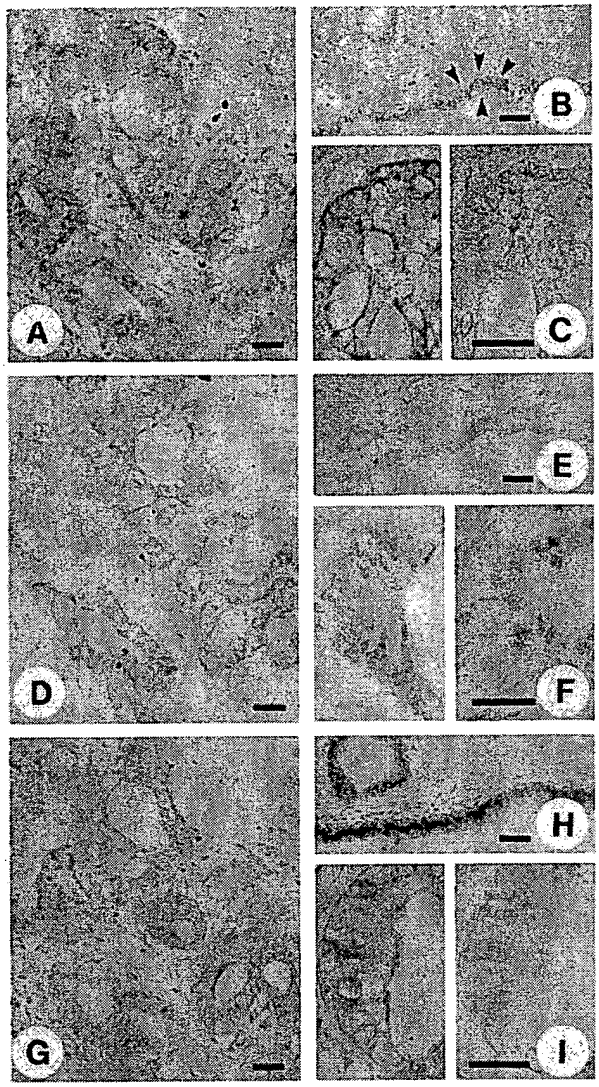


Figure 4. Normal and cancerous pancreatic tissues were subjected to immunostaining using highly specific anti-Id-1 (A-C), anti-Id-2 (D-F), and anti-Id-3 (G-I) antibodies as described in the Methods section. Moderate to strong Id-1 immunoreactivity was present in the cytoplasm of duct-like cancer cells (A and C, left panel). In the normal pancreas there was weak Id-1 immunoreactivity in the ductal cells (B). Preabsorption with the Id-1-specific blocking peptide abolished the Id-1 immunoreactivity (C, right panel). Strong Id-2 immunoreactivity was observed in the cytoplasm of the cancer cells that exhibited duct-like structures (D and F, left panel), whereas in the normal pancreas, there was only weak Id-2 immunoreactivity in the ductal cells (E). Preabsorption with the Id-2-specific blocking peptide abolished the Id-2 immunoreactivity (F, right panel). Moderate to strong Id-3 immunoreactivity was present in the duct-like cancer cells (G and I, left panel). Moderate to strong Id-3 immunoreactivity was also present in the ductal cells of normal pancreatic tissue samples (H). Id-3 immunoreactivity was completely abolished by preabsorption with the Id-3 specific blocking peptide (I, right panel). A, D, and G constitute serial sections of a pancreatic cancer sample, revealing coexpression of the three Id proteins. Scale bars, 25 μm.

Cruz, CA) polyclonal antibodies were used for immunhistochemistry. These affinity-purified rabbit polyclonal antibodies specifically react with Id-1, Id-2, and Id-3, respectively, of human origin, as determined by Western blotting. Paraffin-embedded sections (4 μm) were subjected to immunostaining using the streptavidin-peroxidase technique. Where indicated, immunostaining for all three Id proteins was performed on serial sections. Endogenous peroxidase activity was blocked by incubation for 30 minutes with 0.3% hydrogen peroxide in methanol. Tissue sections were incubated for 15 minutes (23°C) with 10% normal goat serum and then incubated for 16 hours at 4°C with the indicated antibodies in PBS containing 1% bovine serum albumin. Bound antibodies were detected with biotinylated goat anti-rabbit IgG secondary antibodies and streptavidin-peroxidase complex, using diaminobenzidine tetrahydrochloride as the substrate. Sections were counterstained with Mayer's hematoxylin. Preabsorption with Id-1-, Id-2-, or Id-3-specific blocking peptides completely abolished immunoreactivity of the respective primary antibody. The immunohistochemical results were semiquantitatively analyzed as described previously.[29,30] The percentage of positive cancer cells was stratified into four groups: 0, no cancer cells exhibiting immunoreactivity; 1, <33% of the cancer cells exhibiting immunoreactivity; 2, 33 to 67% of the cancer cells exhibiting immunoreactivity; 3 >67% of the cancer cells exhibiting immunoreactivity. The intensity of the immunohistochemical signal was also stratified into four groups: 0, no immunoreactivity; 1, weak immunoreactivity; 2, moderate immunoreactivity; 3, strong immunoreactivity. Finally, the sum of the results of the cell score and the intensity score was calculated. Statistical analysis was performed with SigmaStat software. The rank sum test was used, and $P < 0.05$ was taken as the level of significance.

Results

Northern blot analysis of total RNA isolated from 12 normal pancreatic tissues and 16 pancreatic cancers revealed the presence of the 1.2-kb Id-1 transcript and the 1.6-kb Id2 mRNA transcript in 11 of the 12 normal pancreatic samples, and the 1.3-kb Id-3 mRNA transcript in all normal pancreatic samples (Figure 1A, 2). In the cancer tissues, Id-1 mRNA levels were elevated in 8 of 16 samples, Id-2 mRNA levels were elevated in 9 of these samples, and Id-3 mRNA levels were elevated in 6 of these samples (Figure 1A, 2). Concomitant overexpression of all three Id species was observed in 6 of the cancer samples (38%). In contrast, none of the Id mRNA species were overexpressed in CP by comparison with normal controls (Figure 1B, 2). Densitometric analysis of all of the autoradiograms indicated that there was a 6.5-fold increase ($P < 0.01$) in Id-1 mRNA levels, a fivefold increase ($P < 0.01$) in Id-2 mRNA levels, and a twofold increase ($P = 0.027$) in Id-3 mRNA levels in the pancreatic cancer samples in comparison to normal controls (Figure 2). In contrast, there was no statistically significant difference in the expression levels of Id-1, Id-2, and Id-3, in CP tissues in comparison to the corresponding levels in the normal pancreas (Figure 2).

Next, we assessed the expression of the three Id genes in 5 human pancreatic cancer cell lines by Northern and Western blot analyses. Id-1 mRNA was present at varying levels in all 5 cell lines (Figure 3). ASPC-1, CAPAN-1, MIA-PaCa-2, and PANC-1 expressed moderate to high levels of Id-1 mRNA, whereas COLO-357 cells

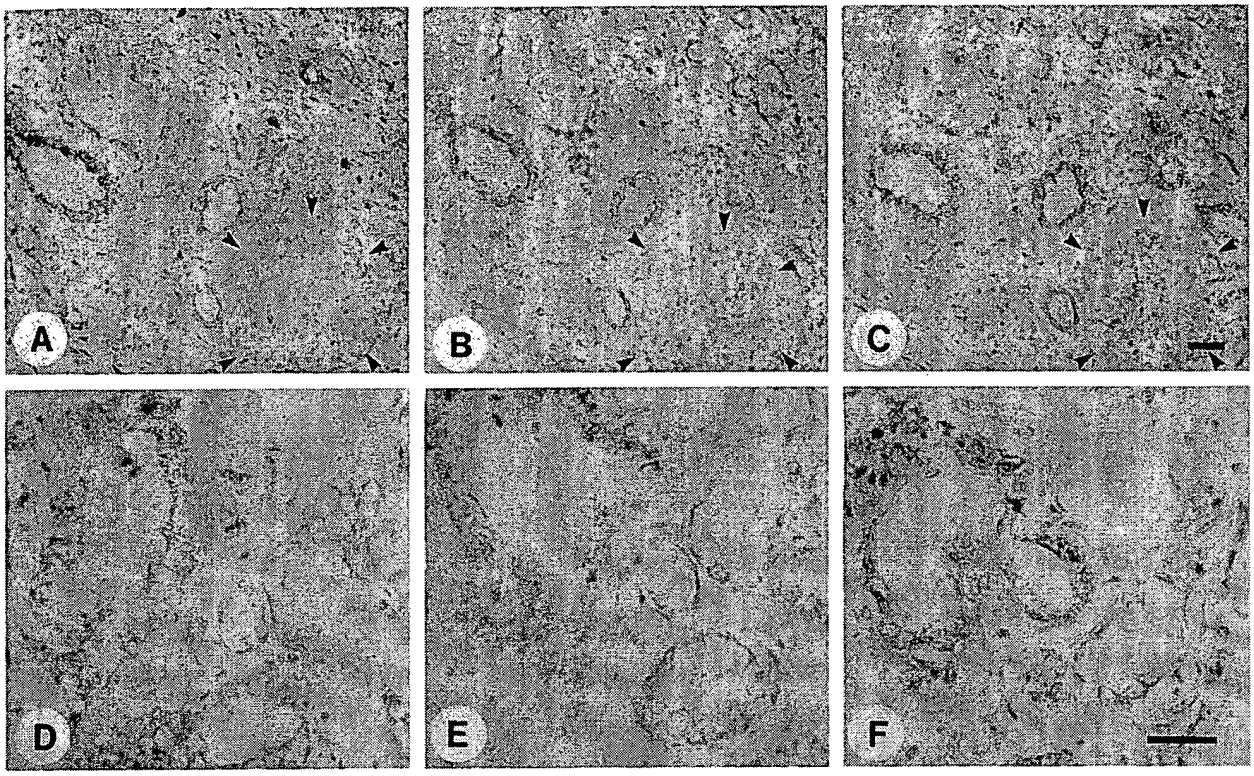


Figure 5. Immunohistochemistry of pancreatic cancer and dysplastic ducts in CP tissues. In the pancreatic cancer tissues (A-C) there was moderate to strong Id-1 (A), Id-2 (B), and Id-3 (C) immunoreactivity in the ductal cells in the areas adjacent to the cancer cells that exhibited CP-like alterations. Islet cells did not exhibit Id immunoreactivity (outlined by solid arrowheads). In the CP samples, moderate to strong Id-1 (D), Id-2 (E), and Id-3 (F) immunoreactivity was present in the cytoplasm of epithelial cells forming large dysplastic ducts. Scale bar, 25 μm.

expressed relatively low levels of this mRNA moiety. Western blotting with a highly specific anti-Id-1 antibody confirmed the presence of the approximately 14-kd Id-1 protein in the 4 cell lines that expressed high levels of Id-1 mRNA (Figure 3). Furthermore, the three cell lines with the highest Id-1 mRNA expression (CAPAN-1, MIA-PaCa-2, and PANC-1) also exhibited the highest Id-1 protein expression. Variable levels of the 1.6-kb Id-2 mRNA transcript were present in all 5 cell lines. In addition, a minor band of approximately 1.2 kb was visible in COLO-357 and MIA-PaCa-2 cells. Immunoblot analysis with a highly specific anti-Id-2 antibody revealed two bands of approximately 16 and 18 kd at relatively high levels in all of the cell lines with exception of PANC-1 cells, in which the 16-kd band was relatively faint (Figure 3). With the exception of MIA-PaCa-2 cells, there was a good correlation between Id-2 mRNA and protein levels (Figure 3). Id-3 mRNA was present at high levels in MIA-PaCa-2 cells, at moderate levels in COLO-357 cells, and at low levels in PANC-1 cells. Id-3 mRNA was not detectable in ASPC-1 and CAPAN-1 cells (Figure 3). Immunoblot analysis with a highly specific anti-Id-3 antibody revealed an approximately 14-kd band that was most abundant in MIA-PaCa-2 cells, and was also readily apparent in COLO-357 and PANC-1 cells. In contrast, only a faint Id-3 band was seen in ASPC-1 and CAPAN-1 cells. Thus, with the exception of PANC-1 cells, there was a good correlation between Id-3 mRNA and protein levels.

To determine the localization of Id-1, Id-2, and Id-3, immunostaining was carried out using the same highly specific anti-Id antibodies. In the pancreatic cancers, moderate to strong Id-1 immunoreactivity was present in the cancer cells in 9 of 10 randomly selected cancer samples. An example of moderate Id-1 immunoreactivity is shown in Figure 4A, and of strong immunoreactivity in Figure 4C (left panel). In contrast, in the normal pancreas, faint Id-1 immunoreactivity was present only in the ductal cells of pancreatic ducts (Figure 4B, arrowheads). Preabsorption with the Id-1-specific blocking peptide completely abolished the Id-1 immunoreactivity (Figure 4C, right panel). The cancer cells also exhibited strong Id-2 (Figure 4, D and F, left panel) and moderate to strong Id-3 immunoreactivity. An example of moderate Id-3 immunoreactivity is shown in Figure 4G, and of strong immunoreactivity in Figure 4I (left panel). In contrast, only faint Id-2 immunoreactivity was present in the ductal cells in the normal pancreas (Figure 4E), whereas Id-3 immunoreactivity in these cells was more variable and ranged from moderate to occasionally strong (Figure 4H). Islet cells and acinar cells were always devoid of Id immunoreactivity. Preabsorption of the respective antibody with the blocking peptides specific for Id-2 (Figure 4F, right panel) and Id-3 (Figure 4I, right panel) completely abolished immunoreactivity. Analysis of serial pancreatic cancer sections revealed that there was often colocalization of the

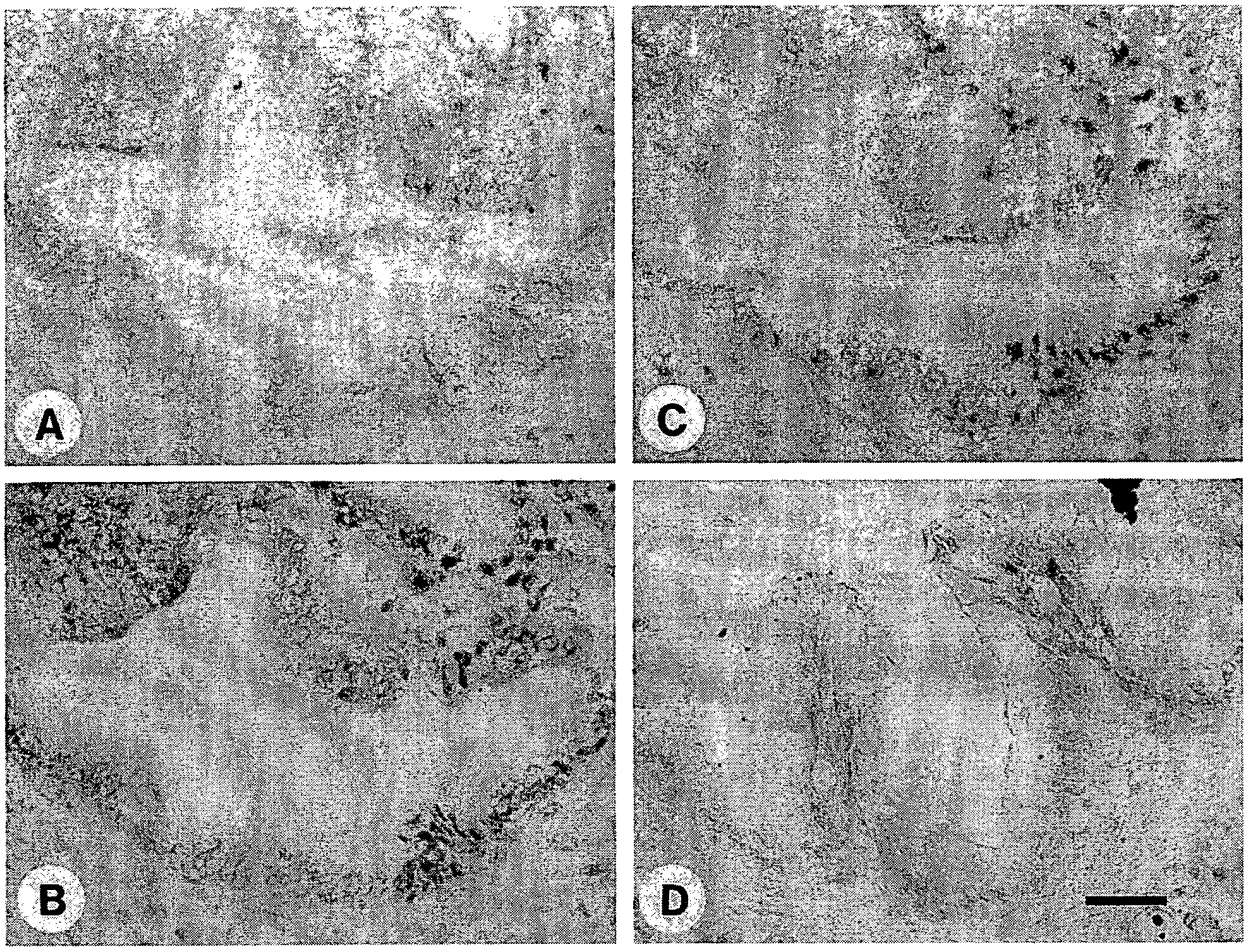


Figure 6. Immunohistochemistry of atypical papillary epithelium in CP tissues. Serial section analysis of some CP samples revealed the presence of large duct-like structures with atypical papillary epithelium. Mild to moderate Id-1 (A) and Id-2 (B) immunoreactivity and weak Id-3 (C) immunoreactivity was present in the cytoplasm of the cells forming these large ducts with papillary structures. Some CP samples also exhibited moderate Id-3 immunoreactivity in these cells (D). Scale bar, 25 μm.

three Id proteins. An example of serial sections from a pancreatic cancer tissue is shown in Figure 4, A, D, and G.

Id-1, Id-2, and Id-3 immunoreactivity was also present at moderate levels in the cytoplasm of ductal cells within CP-like areas adjacent to the cancer cells (Figure 5, A-C). As in the normal pancreas, islet cells (outlined by arrowheads) did not exhibit Id immunoreactivity. In 4 of 9 CP samples, there were foci of ductal cell dysplasia of relatively large interlobular ducts, all of which exhibited moderate to strong Id-1, Id-2, and Id-3 immunoreactivity (Figure 5, D-F). Five of 9 CP samples also contained foci of large ducts exhibiting atypical papillary epithelium. Serial section analysis of one of those CP samples revealed mild to moderate Id-1 and Id-2 immunoreactivity and weak Id-3 immunoreactivity in the cells of these atypical papillary ducts (Figure 6, A-C). In contrast, in some of these CP samples, moderate to strong Id-3 immunoreactivity was also observed (Figure 6D). However, most of the ductal cells forming the typical ductular structures of CP, such as large interlobular ducts and small proliferating ducts, exhibited generally only weak to occasionally moderate Id immunoreactivity (data not shown).

The immunohistochemical data for Id-1, Id-2, and Id-3 are summarized in Table 1. In the case of Id-1 and Id-2, the cancer cells as well as the dysplastic and atypical papillary ducts in CP exhibited a significantly higher score than the ductal cells in the normal pancreas. In contrast, due to the marked variability in Id-3 immunostaining in the normal pancreas, the differences between normal and cancer cells and normal and dysplastic cells did not achieve statistical significance.

Discussion

Id proteins constitute a family of HLH transcription factors that are important regulators of cellular differentiation and proliferation.[1,2] To date, four members of the human Id family have been identified.[1,10–12] Their expression is enhanced during cellular proliferation and in response to mitogenic stimuli,[19,31] and overexpression of Id genes inhibits differentiation and/or enhances proliferation in several different cell types.[15,32–34] The forced expression of Id-1 in mouse small intestinal epithelium results in

Table 1. Histological Scoring

		Id-1	Id-2	Id-3
Normal (n = 6)	Ductal cells	2.0 ± 0.4	2.3 ± 0.2	2.5 ± 0.9
Cancer (n = 10)	Cancer cells	4.5* ± 0.5	5.2§ ± 0.3	4.5 ± 0.6
CP (n = 9)	Typical CP lesions (n = 9)	2.7 ± 0.5	3.1 ± 0.6	3.4 ± 0.7
	Dysplastic ducts (n = 4)	5.3† ± 0.2	5.8‡ ± 0.2	5.3 ± 0.4
	Atypical papillary ducts (n = 5)	4.4‡ ± 0.2	5.2‡ ± 0.2	5.0 ± 0.4

Scoring of the histological specimens was performed as described in the Patients and Methods section. Values are the means ± SD of the number of samples indicated in parenthesis. P values are based on comparisons with the respective controls in the normal samples.
*, $P < 0.02$; †$P < 0.01$; ‡$P = 0.004$; §$P = 0.001$.

adenoma formation in these animals.[35] The growth-promoting effects of Id genes are thought to occur through several mechanisms. For example, Id-2 can bind to members of the pRB tumor suppressor family, thus blocking their growth-suppressing activity,[20,21] and Id-1 and Id-2 can antagonize the bHLH-mediated activation of known inhibitors of cell cycle progression such as the cyclin-dependent kinase inhibitor p21.[23]

In the present study, we determined by Northern blot analysis that a significant percentage of human pancreatic cancers expressed increased Id-1, Id-2, and Id-3 mRNA levels. Increased expression was most evident for Id-1 (6.5-fold) and Id-2 (fivefold). In contrast, Id-3 mRNA levels were only twofold increased in the cancer samples, partly because this mRNA was present at relatively high levels in the normal pancreas. Immunhistochemical analysis confirmed the presence of Id-1, Id-2, and Id-3 in the cancer cells within the tumor mass, whereas in the normal pancreas faint Id-1 and Id-2 immunoreactivity and moderate to occasionally strong Id-3 immunoreactivity was present in some ductal cells. Pancreatic acinar and islet cells in the normal pancreas were devoid of Id-1, Id-2, and Id-3 immunoreactivity. In the cancer samples, all three Id proteins often colocalized in the cancer cells. Coexpression of all three Id genes was also observed in cultured pancreatic cancer cell lines, which often exhibited a close correlation between Id mRNA and protein expression. However, in MIA-PaCa-2 there was a divergence of Id-2 mRNA and protein levels, and in PANC-1 cells, Id-3 mRNA levels did not correlate well with Id-3 protein expression. These observations suggest that in these cells, the half-life of either Id mRNA or Id protein may be altered by comparison with the other cell lines. Interestingly, Id-2 immunoblotting revealed two closely spaced bands of approximately 16 and 18 kd in 4 of 5 cell lines. In view of the fact that two possible initiation codons have been reported for the Id-2 gene,[36] our observation raises the possibility that the two Id-2-immunoreactive bands may represent separate translation products of the Id-2 gene.

Pancreatic cancers often harbor p53 tumor suppressor gene mutations[37] and exhibit alterations in apoptosis pathways. Thus, these cancers often exhibit increased expression of anti-apoptotic proteins such as Bcl-2[38] and abnormal resistance to Fas-ligand-mediated apoptosis.[39] It has been shown recently that forced constitutive expression of Id genes together with the expression of anti-apoptotic genes such as Bcl-2 or $BclX_L$ can result in malignant transformation of human fibroblasts,[11] raising the possibility that the enhanced Id expression in pancreatic cancers together with increased expression of anti-apoptotic genes may contribute to the malignant potential of pancreatic cancer cells *in vivo*.

In the CP tissues there was no significant increase in Id-1, Id-2, and Id-3 mRNA levels in comparison to the normal pancreas. Immunohistochemical analysis of pancreatic cancer samples revealed colocalization of weak to moderate Id-1, Id-2, and Id-3 immunoreactivity in proliferating ductal cells in the CP-like regions adjacent to the cancer cells, indicating that Id expression was not restricted to the cancer cells. Similarly, analysis of CP samples indicated weak Id-1, Id-2, and Id-3 immunoreactivity in the cells of small proliferating ducts and large ducts without dysplastic changes. In general, there was a correlation between weak immunoreactivity and low Id mRNA levels. However, in samples that harbored large ducts with papillary structures there was moderate Id immunoreactivity, and in the cells forming dysplastic ducts there was moderate to strong Id immunoreactivity. In these CP samples, Id mRNA levels were relatively higher than in the CP samples that were devoid of these histological changes. Overall, however, increased Id expression, most notably of Id-1 and Id-2, distinguished a subgroup of pancreatic cancers from CP (Table 1).

Epidemiological studies have shown that the risk of developing pancreatic cancer is increased up to 16-fold in patients with pre-existing CP in comparison to the general population.[40] The mechanisms that contribute to neoplastic transformation in CP are not known. Although there is no established tumor progression model for pancreatic cancer, such as the adenoma-carcinoma sequence of colorectal carcinoma,[41] it is generally accepted that K-*ras* and p16 mutations occur relatively early in pancreatic carcinogenesis, whereas p53 mutations occur late in this process.[37,41–43] Increased Id expression may contribute to malignant transformation of cultured cell lines *in vitro*[11] and has been linked to cell invasion in a murine mammary epithelial cell line.[44] In view of the current findings that Id-1, Id-2, and Id-3 are overexpressed in pancreatic cancer and in dysplastic/metaplastic ducts in CP, these observations raise the possibility that elevated levels of Id-1, Id-2, and, to a lesser extent, Id-3 may represent relatively early markers of pancreatic malignant transformation and may contribute to the pathobiology of pancreatic cancer.

References

1. Jan YN, Jan LY: HLH proteins, fly neurogenesis, and vertebrate myogenesis. Cell 1993, 75:827–830
2. Olson EN, Klein WH: bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. Genes Dev 1994, 8:1–8
3. Begley CG, Aplan PD, Denning SM, Haynes BF, Waldmann TA, Kirsch IR: The gene SCL is expressed during early hematopoiesis and encodes a differentiation-related DNA-binding motif. Proc Natl Acad Sci USA 1989, 86:10128–10132
4. Johnson JE, Birren SJ, Anderson DJ: Two rat homologues of Drosophila achaete-scute specifically expressed in neuronal precursors. Nature 1990, 346:858–861
5. Weintraub H: The MyoD family and myogenesis: redundancy, networks, and thresholds. Cell 1993, 75:1241–1244
6. Hu JS, Olson EN, Kingston RE: HEB, a helix-loop-helix protein related to E2A, and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. Mol Cell Biol 1992, 12:1031–1042
7. Langlands K, Yin X, Anand G, Prochownik EV: Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. J Biol Chem 1997, 272:19785–19793
8. Murre C, Bain G, van Dijk MA, Engel I, Furnari BA, Massari ME, Matthews JR, Quong MW, Rivera RR, Stuiver MH: Structure and function of helix-loop-helix proteins. Biochim Biophys Acta 1994, 1218:129–135
9. Murre C, McCaw PS, Vaessin H, Caudy M, Jan LY, Jan YN, Cabrera CV, Buskin JN, Hauschka SD, Lassar AB, Baltimore D: Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell 1989, 58: 537–544
10. Benezra R, Davis RL, Lockshon D, Turner DL, Weintraub H: The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell 1990, 61:49–59
11. Norton JD, Atherton GT: Coupling of cell growth control and apoptosis functions of Id proteins. Mol Cell Biol 1998, 18:2371–2381
12. Norton JD, Deed RW, Craggs G, Sablitzky F: Id helix-loop-helix proteins in cell growth and differentiation. Trends Cell Biol 1998, 8:58–65
13. Christy BA, Sanders LK, Lau LF, Copeland NG, Jenkins NA, Nathans D: An Id-related helix-loop-helix protein encoded by a growth factor-inducible gene. Proc Natl Acad Sci USA 1991, 88:1815–1819
14. Kawaguchi N, DeLuca HF, Noda M: Id gene expression and its suppression by 1,25-dihydroxyvitamin D3 in rat osteoblastic osteosarcoma cells. Proc Natl Acad Sci USA 1992, 89:4569–4572
15. Kreider BL, Benezra R, Rovera G, Kadesch T: Inhibition of myeloid differentiation by the helix-loop-helix protein Id. Science 1992, 255: 1700–1702
16. Le Jossic C, Ilyin GP, Loyer P, Glaise D, Cariou S, Guguen-Guillouzo C: Expression of helix-loop-helix factor Id-1 is dependent on the hepatocyte proliferation and differentiation status in rat liver and in primary culture. Cancer Res 1994, 54:6065–6068
17. Sun XH, Copeland NG, Jenkins NA, Baltimore D: Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins. Mol Cell Biol 1991, 11:5603–5611
18. Wilson RB, Kiledjian M, Shen CP, Benezra R, Zwollo P, Dymecki SM, Desiderio SV, Kadesch T: Repression of immunoglobulin enhancers by the helix-loop-helix protein Id: implications for B-lymphoid-cell development. Mol Cell Biol 1991, 11:6185–6191
19. Hara E, Yamaguchi T, Nojima H, Ide T, Campisi J, Okayama H, Oda K: Id-related genes encoding helix-loop-helix proteins are required for G1 progression and are repressed in senescent human fibroblasts. J Biol Chem 1994, 269:2139–2145
20. Iavarone A, Garg P, Lasorella A, Hsu J, Israel MA: The helix-loop-helix protein Id-2 enhances cell proliferation and binds to the retinoblastoma protein. Genes Dev 1994, 8:1270–1284
21. Lasorella A, Iavarone A, Israel MA: Id2 specifically alters regulation of the cell cycle by tumor suppressor proteins. Mol Cell Biol 1996, 16:2570–2578
22. Peverali FA, Ramqvist T, Saffrich R, Pepperkok R, Barone MV, Philipson L: Regulation of G1 progression by E2A and Id helix-loop-helix proteins. EMBO J 1994, 13:4291–4301
23. Prabhu S, Ignatova A, Park ST, Sun XH: Regulation of the expression of cyclin-dependent kinase inhibitor p21 by E2A and Id proteins. Mol Cell Biol 1997, 17:5888–5896
24. Warshaw AL, Fernandez-del Castillo C: Pancreatic carcinoma. N Engl J Med 1992, 326:455–465
25. Korc M: Role of growth factors in pancreatic cancer. Surg Oncol Clin North Am 1998, 7:25–41
26. Kleeff J, Ishiwata T, Friess H, Büchler MW, Israel MA, Korc M: The helix-loop-helix protein Id2 is overexpressed in human pancreatic cancer. Cancer Res 1998, 58:3769–3772
27. Oertel JE, Heffes CS, Oertel YC: Pancreas. Diagnostic Surgical Pathology. Edited by SS Sternberg. New York, Raven Press, 1989, pp 1057–1093
28. Korc M, Chandrasekar B, Yamanaka Y, Friess H, Büchler MW, Beger HG: Overexpression of the epidermal growth factor receptor in human pancreatic cancer is associated with concomitant increase in the levels of epidermal growth factor and transforming growth factor α. J Clin Invest 1992, 90:1352–1360
29. Saeki T, Stromberg K, Qi CF, Gullick WJ, Tahara E, Normanno N, Ciardiello F, Kenney N, Johmson GR, Salomon DS: Differential immunohistochemical detection of amphiregulin and cripto in human normal colon and colorectal tumors. Cancer Res 1992, 52:3467–3473
30. Cantero D, Friess H, Deflorin J, Zimmermann A, Bründler MA, Riesle E, Korc M, Büchler MW: Enhanced expression of urokinase plasminogen activator and its receptor in pancreatic carcinoma. Br J Cancer 1997, 75:388–395
31. Desprez PY, Hara E, Bissell MJ, Campisi J: Suppression of mammary epithelial cell differentiation by the helix-loop-helix protein Id-1. Mol Cell Biol 1995, 15:3398–3404
32. Shoji W, Yamamoto T, Obinata M: The helix-loop-helix protein Id inhibits differentiation of murine erythroleukemia cells. J Biol Chem 1994, 269:5078–5084
33. Cross JC, Flannery ML, Blanar MA, Steingrimsson E, Jenkins NA, Copeland NG, Rutter WJ, Werb Z: Hxt encodes a basic helix-loop-helix transcription factor that regulates trophoblast cell development. Development 1995, 121:2513–2523
34. Sun XH: Constitutive expression of the Id1 gene impairs mouse B cell development. Cell 1994, 79:893–900
35. Wice BM, Gordon JI: Forced expression of Id-1 in the adult mouse small intestinal epithelium is associated with development of adenomas. J Biol Chem 1998, 273:25310–25319
36. Barone MV, Pepperkok R, Peverali FA, Philipson L: Id proteins control growth induction in mammalian cells. Proc Natl Acad Sci USA 1994, 91:4985–4988
37. Barton CM, Staddon SL, Hughes CM, Hall, PA, O'Sullivan C, Kloppel G, Theis B, Russell RC, Neoptolmos J, Williamson RCN, Lane DP, Lemoine NR: Abnormalities of the p53 tumour suppressor gene in human pancreatic cancer. Br J Cancer 1991, 64:1076–1082
38. Ohshio G, Suwa H, Imamura T, Yamaki K, Tanaka T, Hashimoto Y, Imamura M: An immunohistochemical study of bcl-2 and p53 protein expression in pancreatic carcinomas. Scand J Gastroenterol 1998, 33:535–539
39. Ungefroren H, Voss M, Jansen M, Roeder C, Henne-Bruns D, Kremer B, Kalthoff H: Human pancreatic adenocarcinomas express Fas and Fas ligand yet are resistant to Fas-mediated apoptosis. Cancer Res 1998, 58:1741–1749
40. Niederau C, Niederau MC, Heintges T, Lüthen R: Epidemiology: relation between chronic pancreaitis and pancreatic carcinoma. Cancer of the Pancreas. Edited by HG Beger, MW Büchler, MH Schoenberg. Ulm, Germany, Universitätsverlag Ulm GmbH, 1996, pp 6–9
41. Moskaluk CA, Kern SE: Molecular genetics of pancreatic carcinoma. Pancreatic Cancer: Pathogenesis, Diagnosis, and Treatment. Edited by HA Reber. Totowa, NJ, Humana Press, 1998, pp 3–20
42. Moskaluk CA, Hruban RH, Kern SE: p16 and K-ras gene mutations in the intraductal precursors of human pancreatic adenocarcinoma. Cancer Res 1997, 57:2140–2143
43. Tada M, Ohashi M, Shiratori Y, Okudaira T, Komatsu Y, Kawabe T, Yoshida H, Machinami R, Kishi K, Omata M: Analysis of K-ras gene mutation in hyperplastic duct cells of the pancreas without pancreatic disease. Gastroenterology 1996, 110:227–231
44. Desprez PY, Lin CQ, Thomasset N, Sympson CJ, Bissell MJ, Campisi J: A novel pathway for mammary epithelial cell invasion induced by the helix-loop-helix protein Id-1. Mol Cell Biol 1998, 18:4577–4588

American Journal of Pathology, Vol. 161, No. 4, October 2002

Distinct and Complementary Information Provided by Use of Tissue and DNA Microarrays in the Study of Breast Tumor Markers

Christophe Ginestier,*
Emmanuelle Charafe-Jauffret,*†
François Bertucci,*‡ François Eisinger,§
Jeannine Geneix,* Didier Bechlian,*
Nathalie Conte,* José Adélaïde,* Yves Toiron,*†
Catherine Nguyen,¶ Patrice Viens,‡
Marie-Joelle Mozziconacci,*† Rémi Houlgatte,¶
Daniel Birnbaum,* and Jocelyne Jacquemier*†

From the Département d'Oncologie Moléculaire, Institut Paoli-Calmettes and Institut National de la Santé et de la Recherche Medical U119, IFR57, Marseille; the Departements de Biopathologie† and de Dépistage et Prévention,§ Institut Paoli-Calmettes, Marseille; the Département d'Oncologie Médicale,‡ Institut Paoli-Calmettes, Université de la Méditerranée, Marseille; and the Laboratoire Technologies Avancées pour le Génome et la Clinique,¶ Centre d'Immunologie de Marseille-Luminy, Luminy, Marseille, France*

Emerging high-throughput screening technologies are rapidly providing opportunities to identify new diagnostic and prognostic markers and new therapeutic targets in human cancer. Currently, cDNA arrays allow the quantitative measurement of thousands of mRNA expression levels simultaneously. Validation of this tool in hospital settings can be done on large series of archival paraffin-embedded tumor samples using the new technique of tissue microarray. On a series of 55 clinically and pathologically homogeneous breast tumors, we compared for 15 molecules with a proven or suspected role in breast cancer, the mRNA expression levels measured by cDNA array analysis with protein expression levels obtained using tumor tissue microarrays. The validity of cDNA array and tissue microarray data were first verified by comparison with quantitative reverse transcriptase-polymerase chain reaction measurements and immunohistochemistry on full tissue sections, respectively. We found a good correlation between cDNA and tissue array analyses in one-third of the 15 molecules, and no correlation in the remaining two-thirds. Furthermore, protein but not RNA levels may have prognostic value; this was the case for MUC1 protein, which was studied further using a tissue microarray containing ~600 tumor samples. For THBS1 had prognostic value. Thus, differences extended to clinical prognostic information obtained by the two methods underlining their complementarity and the need for a global molecular analysis of tumors at both the RNA and protein levels. *(Am J Pathol 2002, 161:1223–1233)*

The development of genomic, technological, and bioinformatic tools have allowed progress in cancer research. DNA arrays are currently the most used of the new high-throughput methods to analyze the molecular complexity of tumors. Several studies have showed their potential in many types of human cancers.[1–4] Even if the clinical benefits for patients remain to be demonstrated, the first results are very encouraging. DNA arrays-based gene expression profiles are improving our understanding of the disease as well as tumor taxonomy by identifying new diagnostic or prognostic subclasses unrecognized by usual parameters. They are expected to lead to the discovery of new potential therapeutic targets, to accurate predictions of survival and response to a given treatment, and eventually to the delivery of a therapy appropriate to each individual patient.

Once a potential marker is identified by this technique, an important next step is its validation and introduction in routine tests in hospital settings.[5,6] There, cDNA arrays are not the method of choice because they are still expensive, time-consuming, complex, and require frozen material not always available. Validation studies have been done traditionally by immunohistochemistry (IHC) on paraffin-embedded tissues allowing analysis of many archived samples with a long follow-up. Until recently, pathologists examined sections of tumor slide by slide. Today, the recently developed tissue microarray (TMA) technology[7–9] allows the simultaneous analysis of thousands of tumor samples arrayed onto glass slides. This may facilitate the search for correlations between

Supported by INSERM, Institut Paoli-Calmettes, and grants from l'Association pour la Recherche sur le Cancer and la Ligue Nationale contre le Cancer.

ECJ and FB contributed equally to this work.

Accepted for publication June 20, 2002.

Address reprint requests to Daniel Birnbaum, U119 INSERM, Institut

molecular alterations and the histoclinical features of the tumors.

In a recent cDNA array-based, prognosis-oriented study of 55 localized breast carcinoma samples,[10] we identified two clusters of discriminator genes (named I and II) the differential expression of which allowed to distinguish subclasses of tumors with significantly different clinical outcome after adjuvant chemotherapy. The aim of the present study was to validate some of these data using TMAs and to evaluate the interest and limitations of this technology as a validation tool. Cylinders from the same 55 tumors were arrayed in a specific tissue-microarray and studied by IHC using antibodies directed against proteins encoded by some of our discriminator genes.

Materials and Methods

Mammary Carcinoma Cell Lines

Nine established mammary carcinoma cell lines were used as positive controls for expression of various genes or proteins. They included: BT-474, MCF-7, MCF-10F, MDA-MB-157, MDA-MB-175, MDA-MB-231, MDA-MB-453, BrCa-MZ-02,[11] and HBL-100. All cell lines are derived from carcinomas except HBL-100 and MCF-10F. They were obtained from the American Type Culture Collection, Rockville, MD (*http://www.atcc.org/*) and grown using the recommended culture conditions.

Breast Tumor Samples and Characteristics of Patients

Tumor samples were obtained from 55 women treated at the Institut Paoli-Calmettes. Inclusion criteria were: 1) localized breast cancer treated with adjuvant anthracyclin-based chemotherapy in addition to loco-regional treatment; 2) tumor material quickly macrodissected and frozen in liquid nitrogen and stored at −160°C; and 3) patient follow-up of 48 months or more after diagnosis. In addition to the axillary lymph node status, four poor prognosis criteria were used to determine whether adjuvant chemotherapy should be administered: patient age less than 40 years, pathological tumor size greater than 20 mm, Scarff-Bloom-Richardson grade equal to 3, and negative estrogen receptor (ER) status as evaluated by IHC with a positivity cutoff value of 1%. Women who received chemotherapy were those with either node-positive tumors or node-negative tumors and one of the poor prognosis criteria if nonmenopausal or two criteria if menopausal. All tumor sections were *de novo* reviewed by a pathologist (JJ) before analysis; all samples contained more than 50% tumor cells. Tumors were infiltrating adenocarcinomas including, according to the World Health Organization histological typing, 42 ductal, 5 lobular, 5 mixed, and 3 medullary carcinomas.

A second series of breast tumors was analyzed. It was constituted by 592 localized forms of breast cancer collected between 1987 and 1999 (median follow-up 48 months) and frozen in liquid nitrogen (the 55 tumors previously described were included in this array). There were 401 ductal, 77 lobular, 40 mixed, 4 medullary carcinomas, and 70 other histological types. A total of 297 tumors were node positive and 450 were positive for ER.

Extraction of RNA from Frozen Tissue

Total RNA was extracted from tumor samples by standard methods, as previously described.[12] RNA integrity was controlled on denaturing formaldehyde-agarose gel electrophoresis and Northern blots using a 28S-specific oligonucleotide.

DNA Arrays

DNA arrays were made in our facility (Technologies Avanceés pour le Génome et la Clinique)). Nylon filter preparation with spotted polymerase chain reaction (PCR) products derived from ~1000 selected candidate cancer genes, ^{33}P radioactive hybridization, and data acquisition, normalization, and analysis have been described elsewhere[13,14] and can also be consulted on our web site (*http:/tagc.univ-mrs.fr/pub/Cancer/*).

Reverse Transcription

RNA extracted from frozen tissue was reverse-transcribed in a final volume of 20 μl containing 1× reverse transcriptase (RT)-PCR buffer (Invitrogen Corp., Carlsbad, CA) , 5 mmol/L $MgCl_2$ (Invitrogen), 1 mmol/L dXTP (Roche Diagnostics, Meylan, France), 10 mmol/L dithiothreitol (Invitrogen), 5 μmol/L random hexamers (Roche), 20 U of RNase inhibitor (Promega Biosciences, Madison, WI) , 200 U of superscript reverse transcriptase (Invitrogen), and 1 μg of total RNA (calibration curve points and patient samples). Samples were incubated at 20°C for 10 minutes and 42°C for 45 minutes; reverse transcriptase was inactivated by heating at 99°C for 3 minutes and cooling at 4°C for 5 minutes.

Real-Time Quantitative RT-PCR (RQ-PCR)

RQ-PCR analyses for *ERBB2*, *MUC1*, and *TBP* (TATA box binding protein) mRNA were done using the ABI PRISM 7700 Sequence Detection System instrument and software (Perkin Elmer Applied Biosystems, Foster City, CA). Conditions for the analysis of these markers have been described.[15,16] Primers and probes for the TaqMan system were designed to meet specific criteria by using Primer Express software (Perkin Elmer) and were synthesized by Genset (Genset Olijos, La Jolla, CA, USA) for the primers and by Roche for the probes. The 5′- and 3′-end nucleotides of the probe were labeled with a reporter (FAM, 6-carboxy-fluorescein) and a quencher dye (TAMRA, 6-carboxy-tetramethylrhodamine). The sequences of the PCR primer pairs and fluorogenic probes used for each gene are shown in Table 1. The oligonucleotides are designed

Table 1. Sequences of Oligonucleotide Primers and Probes Used in RQ-PCR Experiments

Gene	Oligonucleotide	Sequence	PCR product size
ERBB2	Forward primer	5′-AGCCGCGAGCACCCAAGT-3′ (exon 1)	147 bp
	Reverse primer	5′-TTGGTGGGCAGGTAGGTGAGTT-3′ (exon 2)	
	Probe	5′-CCTGCCAGTCCCGAGACCCACCT-3′	
MUC1	Forward primer	5′-ACCATCCTATGAGCGAGTACC-3′ (exon 6)	107 bp
	Reverse primer	5′-GTTTCTGCAGGTAATGGTGGC-3′ (exon 7)	
	Probe	5′-CCCATGGGCGCTATGTGCCC-3′	
TBP	Forward primer	5′-CACGAACCACGGCACTGATT-3′	89 bp
	Reverse primer	5′-TTTTCTTGCTGCCAGTCTGGAC-3′	
	Probe	5′-TGTGCACAGGAGCCAAGAGTGAAG-3′	

Bank accession no. M11730, *MUC1* GenBank accession no. JO5581, *TBP* GenBank accession no. X54993. The precise amount of total RNA added to each reaction mix (based on absorbance) and its quality (ie, lack of extensive degradation) are both generally difficult to assess. Therefore, the relative expression level of the gene of interest was computed with respect to the internal standard *TBP* to normalize for variations in the quality of RNA and the amount of input cDNA. Ct (threshold cycle) was used for quantification of the input target number and all experiments were done with duplicates for each data point. All patient samples with a variation >1 Ct for the duplicate were retested. For each experimental sample, the amount of target and endogenous reference was determined from a standard curve. The standard curve was constructed with fivefold serial dilutions of cDNA (1000 ng to 1 ng) from BT-474 (for *ERBB2*) and MCF-7 (for *MUC1*) breast carcinoma cell lines, respectively. The relative target gene expression in a tested sample was normalized using a calibrator sample, ie, the HME1 human primary mammary epithelial cell line (Clontech). The level of expression of the target gene was given by the N-ratio, in which each normalized gene value (*ERBB2*, *MUC1*) was divided by a calibrator normalized gene value (*TBP*).

$$N_{ERBB2} = \frac{ERBB2_{SAMPLE}}{TBP_{SAMPLE}} \Big/ \frac{ERBB2_{CALIBRATOR}}{TBP_{CALIBRATOR}}$$

$$N_{MUC1} = \frac{MUC1_{SAMPLE}}{TBP_{SAMPLE}} \Big/ \frac{MUC1_{CALIBRATOR}}{TBP_{CALIBRATOR}}$$

PCR was done with 1× TaqMan Universal PCR Master Mix (Perkin Elmer), 300 nmol/L of primers, 200 nmol/L of the probe, and 1 μl of each appropriately diluted reverse transcription sample in a 25-μl final reaction mixture. After a 2-minute incubation at 50°C to allow for uracyl N-glycosylate cleavage, AmpliTaq Gold was activated by an incubation for 10 minutes at 95°C. Each of the 40 PCR cycles consisted of 15 seconds of denaturation at 95°C and hybridization of probe and primers for 1 minute at 60°C.

TMA Construction

TMAs were prepared as described[9] with slight modifications. For each tumor, three representative tumor areas were carefully selected from a hematoxylin- and eosin-stained section of a donor block. Core cylinders with a diameter of 0.6 mm each were punched from each of these areas and deposited into a recipient paraffin block

Table 2. List of Proteins Tested by Immunohistochemistry and Characteristics of the Corresponding Antibodies

Protein	Antibody	Origin	Clone	Dilution
Angiogenin (ANG)	Rabbit polyclonal	Santa Cruz Biotechnology	sc-9044	1/20
BCL2	mmab	DAKO	124	1/100
E Cadherin (CDH1)	mmab	Transduction Laboratories	36	1/2000
ERBB2	mmab	Novocastra Laboratories Ltd.	CB 11	1/500
ERBB2	mmab	Oncogene Research Products	3B5	1/500
ERBB2	Rabbit polyclonal	DAKO	AO 485	1/1000
Estrogen receptor (ESR1/ER)	mmab	Novocastra Laboratories Ltd.	6F11	1/60
FGFR1	Rabbit polyclonal	Santa Cruz Biotechnology	sc-121	1/200
GATA3	mmab	Santa Cruz Biotechnology	sc-268	1/100
Ki67	mmab	DAKO	KI-67	1/100
Melan A/MART1 (MLANA)	mmab	DAKO	A103	1/2
MUC1	mmab	Transgen	H23	1/1000
P53	mmab	Immunotech	DO-1	1/4
Progesterone receptor (PR)	mmab	DAKO	PgR 636	1/80
Prolactin receptor (PRLR)	mmab	NeoMarkers	B6.2	1/200
Transforming acidic coiled-coil 1 TACC1	Rabbit polyclonal	Upstate Biotechnology	07-229	1/200
Transforming acidic coiled-coil 2 TACC2	Rabbit polyclonal	Upstate Biotechnology	07-228	1/40
Thrombospondin 1 (THBS1)	mmab	Oncogene Research Products	46.4	1/10

using a specific arraying device (Beecher Instruments, Silver Spring, MD). In addition to tumor tissues, the recipient block also received normal breast tissue and cell line pellets. Five-μm sections of the resulting microarray block were made and used for IHC analysis after transfer to glass slides. Two TMAs were prepared; the first one contained the 55 tumors studied by cDNA arrays (with three cores per sample) and controls, the second one was used for MUC1 study and contained 592 tumor samples (with one core per sample) and controls.

Antibodies and IHC

The characteristics of the antibodies used are listed in Table 2. IHC was performed on 5-μm sections of formalin-embedded tissue specimens. They were deparaffinized in histolemon (Carlo Erba Reagenti, Rodano, Italy) and rehydrated in graded alcohol. Antigen enhancement was done by incubating the sections in target retrieval solution (DAKO, Copenhagen, Denmark) as recommended except for prolactin receptor, in which pretreatment was done with incubation in pepsin (Zymed Laboratories, South San Francisco, CA), for 30 minutes at 37°C, and for MUC1, in which no pretreatment was done. Slides were then transferred to a DAKO autostainer. Staining was done at room temperature as follows: after washes in phosphate buffer, followed by quenching of endogenous peroxidase activity by treatment with 0.1% H_2O_2, slides were first incubated with blocking serum (DAKO) for 10 minutes and then with the affinity-purified antibody for 1 hour. After washes, slides were incubated with biotinylated antibody against rabbit Ig for 20 minutes followed by streptavidin-conjugated peroxidase (DAKO LSAB^{R}2 kit). Diaminobenzidine or 3-amino-9-ethylcarbazole was used as the chromogen, counterstained with hematoxylin, and coverslipped using Aquatex (Merck, Darmstadt, Germany) mounting solution. Slides were evaluated under a light microscope by two pathologists (EC-J, JJ).

Immunoreactivities were classified by estimating the percentage (P) of tumor cells showing characteristic staining (from undetectable level or 0%, to homogeneous staining or 100%) and by estimating the intensity (I) of staining (1, weak staining; 2, moderate staining; or 3, strong staining. The cutoff values were the same for all markers tested. Results were scored by multiplying the percentage of positive cells by the intensity, ie, by the so-called quick score (Q) ($Q = P \times I$; maximum = 300). For Ki67, only the percentage (P) of tumor cells was estimated, because intensity does not vary. Expression levels allowed to group tumors into four categories: negative expression ($Q = 0$ or $P = 0$ for Ki67), weak expression ($0 < Q \leq 120$ or $0 < P < 25$ for Ki67), moderate expression ($120 < Q \leq 210$ or $25 \leq P < 60$ for Ki67) and strong expression ($210 < Q \leq 300$ or $60 \leq P \leq 100$ for Ki67). Because of its prognostic impact the topographical localization of MUC1 was taken into account and expressed in four categories: absence, apical, circumferential membrane, and cytoplasmic

IHC on Full Tissue Sections

To validate the use of TMAs for immunophenotyping, we compared the protein expression levels of ER, progesterone receptor, P53, and BCL2, on full tissue sections and on TMAs for the group of 55 tumors. The data on full sections were compared to the mean of intensities of the three 0.6-mm core biopsies for 47 cases, or of only two core biopsies for 8 cases.

Statistical Analysis

The concordance between RNA expression levels measured by real-time quantitative RT-PCR and cDNA arrays was examined using Spearman's rank correlation. Comparison between IHC data from full sections and TMAs analyses was measured using κ statistics (a κ value >0.7 indicated a strong association). Contingency table analysis was used to analyze the relationship between protein expression obtained by IHC on TMAs and RNA expression obtained with cDNA arrays (total chi-square test). Survival analysis used the Kaplan-Meier method and survival curves were compared using the log-rank test (a P value <0.5 was considered as significant). All P values were two-sided. To assess the relationship between two variables assumed to be related (ie, the co-regulated molecules), simple linear regression analyses were performed using Excel Software (Microsoft). For these tests, (O,O) points were removed; the relationship tested was thus for cases with at least one positive value. Each result is given with: N the sample size, a the slope of the regression line, the P value and $r2$, the coefficient of determination. Thus, for each positive comparison a linear relationship can be determined (eg, $y = 0.8x + 20$ means BCL2 = 0.8ER + 20).

Results

Selection of Molecules

We previously analyzed the mRNA expression profiles of ~1000 selected genes in 55 breast carcinoma samples using home-made cDNA arrays. Tumors were homogeneous with respect to histological and clinical parameters, and all patients had received adjuvant anthracyclin-based chemotherapy. Detailed results are described elsewhere.[10] Briefly, molecular profiling combined with hierarchical clustering allowed the identification, among this set of poor-prognosis localized breast cancers, of new subclasses distinct with respect to overall and metastasis-free survivals. Such a classification resulted from the differential expression of two discriminator gene clusters (named I and II) and was not possible using classical prognostic factors of disease. Cluster I included the *ESR1* gene encoding ER-α. For the present study, we selected 10 of these genes. Interestingly, six of them (*BCL2*, *ERBB2*, *ESR1*, *GATA3*, *MUC1*, *PRLR*) have also been frequently identified as discriminator genes in expression-profiling studies of breast cancer that have addressed the prognosis issue.[4,17–19] These genes were

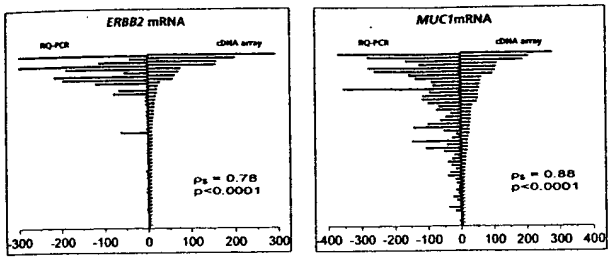


Figure 1. Expression levels of *ERBB2* and *MUC1* mRNA levels measured by cDNA array analysis and real-time quantitative PCR amplification. *ERBB2* and *MUC1* mRNA expression levels measured using cDNA arrays (artificially ×30 for visual effect) (**left**) and real-time quantitative PCR amplification (artificially ×30 for visual effect) (**right**). Results for each tumor (from top to bottom) are represented as **opposite bars**. For *ERBB2*: $\rho_s = 0.78$, $P < 0.0001$; for *MUC1*: $\rho_s = 0.88$, $P < 0.0001$.

thus interesting candidates for further investigation. In addition, other molecules, such as CDH1, Ki67, TP53, progesterone receptor, TACC1,[20] and TACC2, were retained because of a known or suspected role in breast cancer. The selection criteria for all molecules also included availability of a commercial antibody. The complete list of the corresponding proteins tested in the following experiments is given in Table 2.

Validation of cDNA Array Data with RQ-PCR

Our cDNA array analyses regularly included extensive experiments and controls designed to ensure reproducibility and reliability of expression measurements.[1,13,14,21] Nevertheless, we sought to further validate our data by comparing RNA expression levels of two genes, *ERBB2* and *MUC1*, as measured by cDNA array, to those obtained by RQ-PCR.

RNA from 50 of 55 samples (RNA was no longer available for five cases) was reverse-transcribed and PCR amplification of *ERBB2* and *MUC1* cDNA was done using a TaqMan device. For *ERBB2*, 41 tumors displayed mRNA expression levels comparable to normal breast and HME1 control cell line, whereas nine samples (18%) showed overexpression. For *MUC1*, 17 tumors displayed mRNA expression levels comparable to normal breast and HME1 control cell line, whereas 33 samples (66%) showed overexpression. As shown in Figure 1, mRNA expression levels obtained with both methods were highly similar (Spearman test: *ERBB2*, $\rho_s = 0.78$, $P < 0.0001$; *MUC1*, $\rho_s = 0.88$, $P < 0.0001$), further suggesting reliability of our cDNA array data.

TMA Analysis and Validation of Data

To validate our TMA analyses, we compared the expression of four selected proteins (BCL2, ER, P53, progesterone receptor) measured by IHC using either standard full tissue sections or TMAs in the panel of 55 breast tumors. For BCL2 expression, 38 cases (69%) showed positive cytoplasm staining, whereas 17 cases (31%) were negative on analysis of full sections. In comparison, 37 cases (67%) were positive and 18 cases were negative (33%) on TMA. Overall, the concordance was 91% and the nonconcordance was 9% (five cases), resulting in a strong statistical association between the two methods (κ value, 0.78). An even better correlation was found for nuclear expressions of ER, P53, and progesterone receptor, with only 3 discordant cases of 55 for each of them (concordance, 95%; Kappa values, 0.86 to 0.88). This high degree of concordance between IHC on full sections and on TMAs justified further use of TMAs.

Analysis of Breast Tumors Using TMAs

Fifteen proteins, including the four previously cited, were tested by IHC on TMAs. Most of them corresponded to genes we had identified in our two discriminator gene clusters I and II.[10] Other tested molecules corresponded to proteins of interest in breast cancer. Immunostainings were evaluated by the quick score (except for Ki67). Results are shown in Table 3 and Figure 2.

Table 3. Results of IHC Stainings on Tissue Microarrays

	Protein	Location of staining	Normal	Negative	Weak	Moderate	Strong
cDNA array, cluster I gene-encoded	ANG	Cytoplasm + Stroma	(+)	17	19	5	14
	BCL2	Cytoplasm	(+)	23	10	17	5
	ESR1/ER	Nucleus	(+)	22	14	5	14
	GATA3	Nucleus	(+)	26	12	7	10
	MUC1	Cytoplasm	(+)	3	19	8	25
	THBS1	Cytoplasm + Stroma	(+)	16	30	9	0
cDNA array, cluster II gene-encoded	MLANA	Cytoplasm	(+)	22	20	6	7
	PRLR	Membrane	(+)	21	12	7	15
Others	CDH1	Membrane	(+)	6	10	14	25
	ERBB2 (CB 11)	Membrane	(−)	34	14	4	3
	ERBB2 (AO485)	Membrane	(−)	30	11	9	5
	ERBB2 (3B5)	Membrane	(−)	37	8	2	8
	FGFR1	Membrane + Cytoplasm	(+)	20	20	12	3
	Ki67	Nucleus	(+)	4	27	13	11
	P53	Nucleus	(−)	33	12	0	10
	TACC1	Cytoplasm	(+)	21	22	9	3
	TACC2	Cytoplasm	(+)	5	18	13	19

(+) and (−) mean expressed or not in normal breast tissue, respectively.

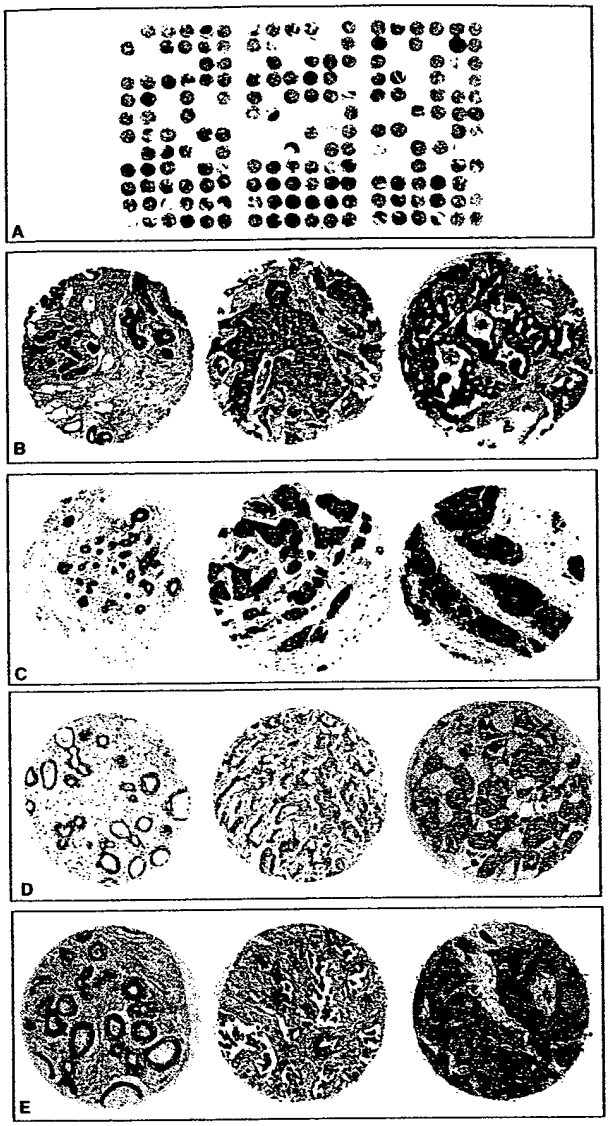


Figure 2. Expression of proteins studied by IHC on TMAs. **A:** H&E staining of a paraffin block section (25 × 30 mm) from the TMA containing 216 arrayed tumor (3 × 55) and control samples. **B:** Anti-angiogenin staining. **C:** Anti-FGFR1 staining. **D:** Anti-GATA3 staining. **E:** Anti-PRLR staining. From **B** to **E**, the first section is from normal breast tissue, the second and third from tumor tissue (the second illustrates a moderate staining whereas the third illustrates a strong staining). Original magnifications, ×50.

Comparison of the Results Obtained by cDNA Arrays and TMAs

Expression levels obtained by IHC on TMA and by cDNA array hybridizations were compared for the 15 molecules. Data from TMA analyses are discontinuous, whereas those obtained by cDNA array analyses are continuous. To facilitate comparisons, we transformed the cDNA array values into discontinuous data. Tumors were then grouped into two or three classes for each method (Table 4). Homogeneous classes were defined for TMA, by grouping tumors with an equivalent staining level (see Table 3). For cDNA arrays, classes were visually defined on examination of the distribution graphs (Figure 3).

Each tumor sample was then placed into one of the three TMA classes and attributed 1, 2, or 3, and into one of the three cDNA array classes and attributed 1, 2, or 3. Table 4 shows the number of samples in each class. Concordance between the two scores was evaluated by a contingency table analysis. A strong concordance was seen for 5 of the 15 comparisons with similar expression levels measured by the two methods: ER, ERBB2, and GATA3 ($P < 0.001$), BCL2 ($P < 0.02$), and TACC1 ($P < 0.05$). No concordance was seen for ANG, CDH1, FGFR1, Ki67, MLANA, MUC1, P53, PRLR, TACC2, and THBS1. Figure 4 shows example of comparative graphs.

Groups of Co-Regulated Molecules

Using cDNA arrays and hierarchical clustering, we had evidenced a co-expression of *ESR1* (encoding ER-α), *BCL2*, and *GATA3* at the mRNA level in breast tumors,[1,10] with a statistically significant correlation between *ESR1* and *GATA3* ($r = 0.73$, $R^2 = 0.53$, $P < 0.0001$). As shown in Figure 5A, the correlation between the three molecules was statistically confirmed at the protein level as measured by IHC on TMA. FGFR1, TACC1, and TACC2 protein levels also varied together but the correlation was weaker (Figure 5B). For each pairwise comparison, with the same number of samples ($n = 55$), we calculated a coefficient of correlation and a P value: BCL2/ER, $r = 0.79$, $R^2 = 0.62$, $P < 0.0001$; GATA3/ER, $r = 0.74$, $R^2 = 0.54$, $P < 0.0001$; TACC1/FGFR1, $r = 0.67$, $R^2 = 0.45$, $P < 0.001$; and TACC2/FGFR1, $r = 0.57$, $R^2 = 0.32$, $P < 0.001$.

Impact on Survival of RNA and Protein Expression Levels

To further estimate the clinical interest of the cDNA array and TMA combined approach, we examined and compared the prognostic information provided by mRNA and protein expression levels for each of the 15 molecules independently. Only 2 of the 15 tested markers showed individual prognostic value. High *THBS1* mRNA levels were associated with a better survival whereas no such correlation was found with protein levels. The opposite was true for MUC1: low levels of MUC1 protein were associated with a better survival, whereas mRNA levels did not correlate with survival (Figure 6). Thus, depending on the marker, clinically relevant information was differently provided by cDNA or TMA technique, suggesting that both analyses are worth performing simultaneously on the same cases.

These results were obtained on a limited number of cases representing a selected population of poor prognosis localized tumors. We sought to confirm the observation on MUC1 on a larger series of cases (Figure 7A). We studied 592 samples (including the 55) arrayed in a second TMA with anti-MUC1 antibody. MUC1 staining in normal cells is either absent or detected in the apical membrane; tumor cells express MUC1 in two abnormal localizations (cytoplasm or circumferential membrane) and a strong cytoplasmic staining is associated with a

Table 4. Comparison of Expression Levels Measured Using Analyses of Tissue Microarrays and cDNA Arrays

Gene	Tissue microarray classes			cDNA array classes			Concordance
	1	2	3	1	2	3	*P* values
ESR1/ER	22 (N)	19 (W +M)	14 (S)	15	22	18	<0.001
BCL2	23 (N)	10 (W)	22 (M+S)	18	37	0	<0.02
P53	33 (N)	22 (W + M+S)	/	46	9	0	NS
GATA3	26 (N)	12 (W)	17 (M+S)	18	25	12	<0.001
PRLR	21 (N)	19 (W +M)	15 (S)	36	12	7	NS
ERBB2 (3B5)	37 (N)	10 (W +M)	8 (S)	46	4	5	<0.001
CDH1	16 (N +W)	14 (M)	25 (S)	42	13	0	NS
TACC2	23 (N +W)	13 (M)	19 (S)	19	18	18	NS
TACC1	21 (N)	22 (W)	12 (M+S)	19	18	18	<0.05
MLANA	22 (N)	20 (W)	13 (M+S)	34	21	0	NS
FGFR1	20 (N)	20 (W)	15 (M+S)	45	10	0	NS
ANG	17 (N)	19 (W)	19 (M+S)	17	30	8	NS
THBS1	16 (N)	30 (W)	9 (M+S)	46	6	8	NS
Ki67	31 (N +W)	24 (M + S)	/	11	33	11	NS
MUC1	22 (N +W)	33 (M + S)	/	39	8	8	NS

N, Negative; W, weak; M, moderate; S, strong; NS, not significant.
Numbers for tissue microarrays are taken from Table 3 and numbers for cDNA arrays are obtained using the method shown in Figure 3.

poor prognosis.[22] For the 55 tumors of the first TMA, the prognostic value of the quantitative quick score was related to a high frequency of abnormal cytoplasmic and circumferential MUC1 localizations (83%) as compared to apical localization and absence (17%). Of the 592 cases of the second TMA, 551 were available for analysis after MUC1 staining: 249 cases (45%) showed apical or no staining, 302 (55%) displayed cytoplasmic or circumferential membrane staining (Figure 7B). In this larger series the quantitative quick score did not have a significant prognostic value. This was because of the fact that the topographical aspect was significantly different from that of the short series with only 55% *versus* 83% of cytoplasmic and circumferential localizations. When considered, qualitative assessment of the staining provided prognostic information; the apical localization and the absence of MUC1 strongly correlated with a better evolution ($P = 0.0154$) (Figure 7C).

Discussion

The recent availability of new high-throughput molecular analyses offers the opportunity to tackle the complexity and the combinatorial nature of breast cancer at the molecular level. Expected applications are a better un-

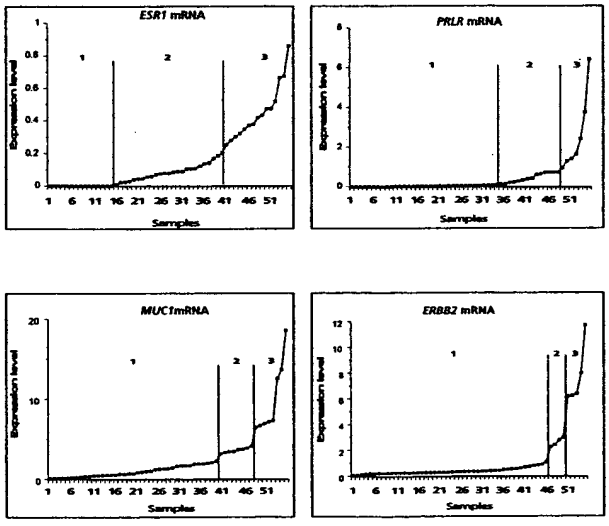


Figure 3. Transformation of continuous cDNA array data into discontinuous data. mRNA expression levels measured by cDNA array are plotted for each sample in an increasing order. For each gene, classes are determined on visual inspection and are separated by **vertical bars** on the graphs. Results for ER-α (*ESR1*), prolactin receptor (*PRLR*), mucin 1 (*MUC1*), and *ERBB2* are shown.

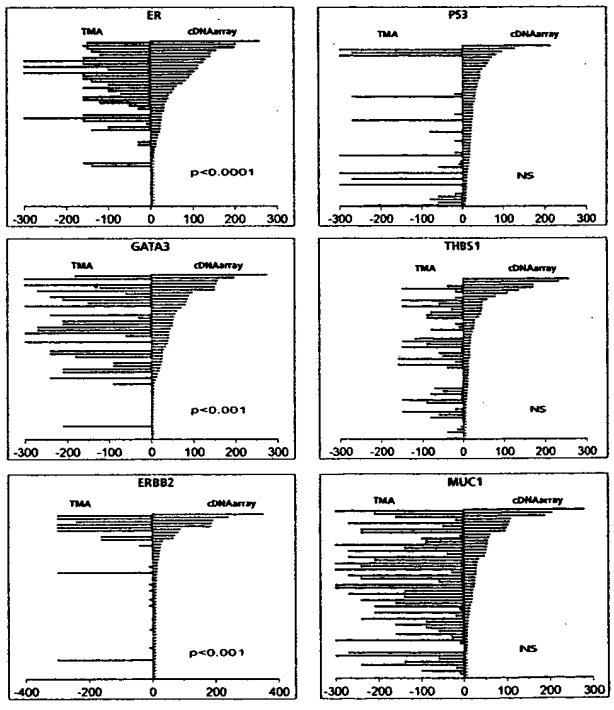


Figure 4. Comparison of data obtained by cDNA array and IHC on TMA. Results for each tumor (from top to bottom) are represented as **opposite bars**, with the value of IHC (quick score) on the **left**, and the value of the cDNA array analyses (artificially ×30 for visual effect) on the **right**. Values for ER, GATA3, and ERBB2 show good correlation between the two methods, whereas values for P53, THBS1, and MUC1 do not show such correlation.

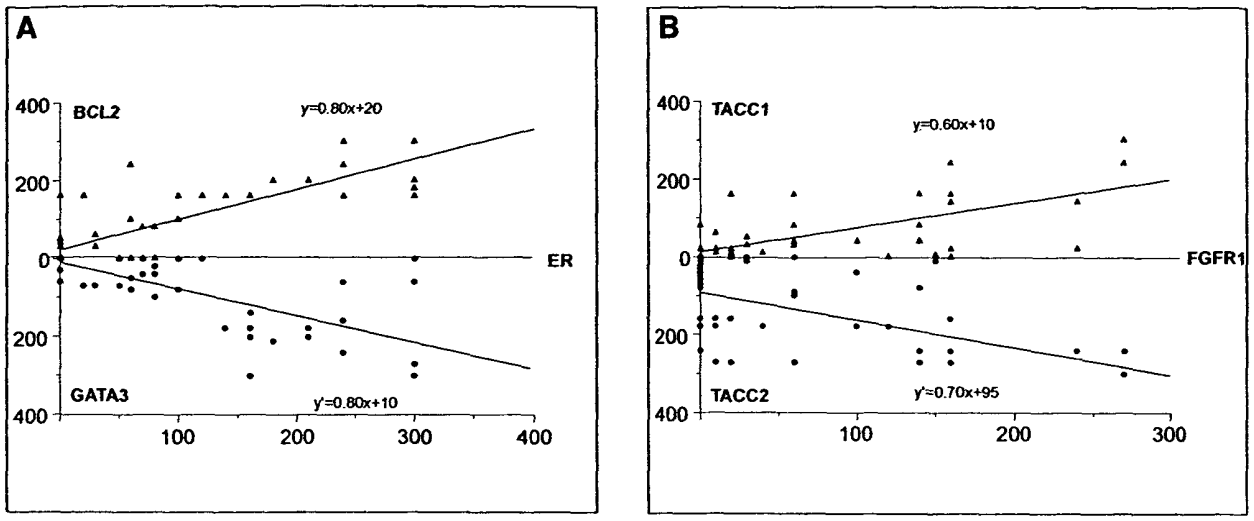


Figure 5. Similar variations in expression levels of two groups of proteins. **A:** The expression levels of ER, BCL2, and GATA3 as measured by IHC on TMAs correlated, as determined by simple linear regression analysis. **B:** Similarly, the expression levels of FGFR1, TACC1, and TACC2 correlated.

derstanding of the disease and the identification of new diagnostic and prognostic markers and therapeutic targets, both needed to improve the management of patients. At the same time it introduces a new challenge for pathologists who, in charge of the first assessment of the tumors, need to know how to optimally use these new methods. The present study directly followed a cDNA array-based analysis of a breast tumor series. The tumor samples were obtained from 55 women with poor prognosis breast cancer treated with adjuvant chemotherapy. Currently such patients have a long-term survival of ~70% and there is a crucial need to identify parameters that might accurately predict the clinical outcome in individual patients. Our study was designed to evaluate the interest and limitations of IHC on TMA as a natural extension of the cDNA array approach in a hospital setting.

We first confronted cDNA array and TMA analyses to other methods, ie, RQ-PCR and conventional IHC, respectively. The good concordance between mRNA expression levels observed by cDNA arrays and RQ-PCR further confirmed the validity of our cDNA array measurements. TMAs allow to screen large series of tumor samples using several archival materials, but their representation of the entire tumor has been questioned. Our degrees of concordance between stainings on full sections and on TMA were in the same range as published studies. Several authors have reported that TMA constructed with three cores per sample (as in our study) are representative of whole tumor specimens.[23–30]

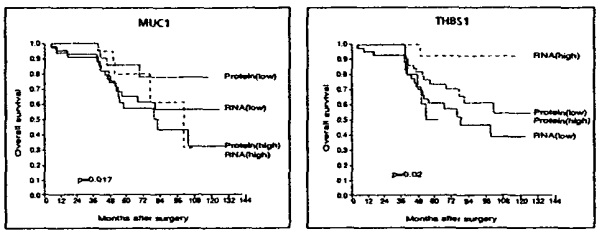


Figure 6. Kaplan-Meier plots of patient overall survival. **Left:** Survival according to *MUC1* mRNA and protein expression levels. **Right:** Survival according to THBS1 mRNA and protein expression levels (labeled high and low). High and low protein levels correspond to strong plus moderate *versus* weak plus negative (see Table 3), respectively, and high and low mRNA levels correspond to classes 2 and 3 *versus* class 1 (see Figure 4), respectively.

As a large-scale validation tool of DNA or RNA data, IHC on TMAs should be interpreted with caution. Indeed, comparison of our cDNA array and TMA data, obtained on the same breast tumor samples, gave different results according to the gene product examined.

For a category of molecules we found important differences between RNA and protein expression levels. This was the case of P53. This discrepancy was rather expected because P53 protein detection is not dependent on mRNA overexpression, but is because of the increased half-life of a mutated protein. In normal cells, P53 protein half-life is short and expression levels are low and undetectable by IHC. In cancer cells, most P53 mutations lead to products that are not ubiquitinated and accumulate in the nuclei where they can then be detected. Other noteworthy cases were MUC1 and THBS1. These differences certainly stem from the fact that different levels of biological information are examined. For many genes, there is little correlation between the abundance of the mRNA transcript with steady-state levels of the encoded protein. Posttranscriptional and posttranslational mechanisms are likely to influence protein expression, thus blurring the correlation between mRNA and protein levels. Proteins encoded by very low levels of RNA, ie, below the detection level of cDNA arrays, can be detected by IHC because of increased protein stability (eg, the case of P53) or high sensitivity of the antibody, and reciprocally, elevated levels of RNA may produce only little amounts of detectable proteins. Special calibration of the antibody aimed to detect only a certain level of protein is another limitation. The chosen antibody may also detect only certain forms of a protein that do not correspond to the cDNA spotted on the DNA array, because of alternative splicings of mRNA for example. This particularly can explain the difference observed between THBS1 mRNA and protein levels, and conse-

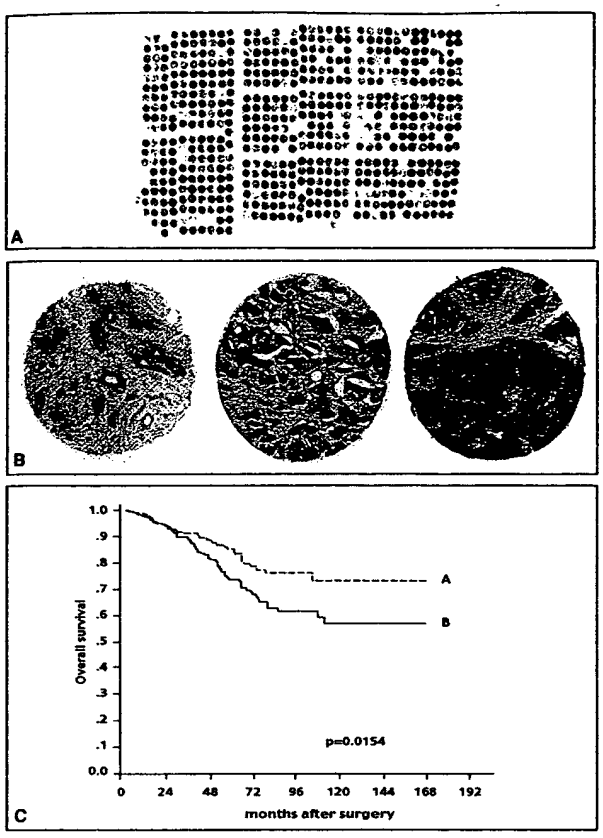


Figure 7. Expression of MUC1 protein studied by IHC on a tissue-microarray. **A:** H&E staining of a paraffin block section (25 × 30 mm) from the TMA containing 647 arrayed samples, including 592 tumors and 55 controls. **B:** MUC1 staining: normal breast tissue (**left**), apical (**middle**), and cytoplasmic (**right**) staining in tumors. **C:** Kaplan-Meier plot of patient overall survival: survival differs significantly according to MUC1 protein localization. **A:** Absence of staining or apical localization; **B:** cytoplasmic or circumferential membrane localization.

quently their different prognostic impact.[31] Finally, distinct areas of a heterogeneous tumor may be submitted to RNA and protein analyses.

Conversely, we observed an excellent correlation between RNA and protein levels in one-third of the tested molecules. This was the case for ERBB2, despite the fact that its corresponding antibody is calibrated to detect only overexpression. Among the other molecules with correlated mRNA and protein expression levels were ER, GATA3, and BCL2. We and others had shown that the mRNA levels of the three genes covaried in cDNA array analyses.[1,10,32] Here we were able to confirm this co-expression at the protein level. This group of co-regulated genes and proteins may be linked to the hormonal control of the mammary gland. Such identification is important for a better understanding of gene and protein networks that operate in cancer cells; it may lead to the discovery of new molecules to be targeted to block or stimulate a metabolic pathway or function; it may also provide a prognostic information clinically more relevant than that of isolated markers because it better reflects the functional status of a pathway such as the estrogen pathway of breast tumors.

Several studies have shown the interest of TMA studies in cancer research to extend cDNA array data.[33] A pioneering analysis was conducted by Moch and colleagues;[8] after the identification of vimentin as overexpressed in a renal cancer cell line using cDNA arrays, the authors extended this result to the protein level on a series of 532 tumor specimens arrayed onto a renal cancer TMA. Using TMA of bladder tumors containing 2317 specimens from 1842 patients, Richter and colleagues[9] found a positive correlation between *CCNE* gene amplification measured by fluorescence *in situ* hybridization and cyclin-E protein overexpression measured by IHC. The combination of cDNA array and TMA allowed the identification of IGFBP2 and HSP27,[34] hepsin,[2] and AMACR[30] as significantly overexpressed in prostate cancer, suggesting their putative diagnostic interest. IGFBP2 was also found as a marker of poor prognosis in a series of 418 brain tumors arrayed onto a TMA.[35] A similar study showed the overexpression of the WT protein in ovarian cancer.[36] The expression level of PKCβ was measured by IHC on a B-cell lymphoma TMA to validate cDNA array data.[3] In breast carcinomas, Hedenfalk and colleagues[37] showed that, like mRNA levels, cyclin D1 protein levels were differentially distributed among BRCA1 and BRCA2 hereditary tumors. All these studies showed a good correlation between the two techniques of investigation, but were limited to the analysis of a single highly selected marker and were not, with few exceptions, conducted on the same samples. Our present study is the first deliberate comparative analysis of cDNA and TMAs. It shows a correlation between the two techniques for one-third of the selected markers and the absence of correlation for the other two-thirds.

These discrepancies deserve two commentaries. First, given the flurry of encouraging data associated with the rapidly emerging cDNA array technology, it is paramount to determine to what extent changes in mRNA expression are accompanied or not by similar changes at the protein level. In some cases, the differences may be eliminated by a number of experimental precautions, such as selection of biopsy cores and antibodies, but in other cases, they will remain. If protein levels of a target molecule, or a group of molecules, correlate with its selection by cDNA array, IHC on TMA offers a powerful tool to quickly evaluate the clinical relevance of differentially expressed genes. But if they do not correlate, the cDNA array and TMA results must be considered independently because each can provide distinct information.

Second, even if the intrinsic prognostic power of cDNA array data and clustering analyses derives from the combined expression of several genes, and not from an individual gene, it may be interesting for routine clinical application to test each of these genes as a candidate marker and to determine how its expression may alone distinguish the tumor classes. The main interest of TMA lies in the possibility to test large series of tumor samples with individual markers. In our series of samples, we observed that mRNA, but not protein expression levels of *THBS1* had prognostic value, suggesting that they play an important role in the discriminator power of the cDNA array gene cluster. In contrast, for MUC1, as seen earli-

er,[38] low levels of protein were associated with a better prognosis, which was not the case for mRNA; IHC further allowed a qualitative appreciation of the protein localization, which happened to be crucial information for prognosis when an unselected population was studied.

In the period of validation studies that has now begun, for which retrospective IHC studies on archival paraffin-embedded material are required,[6] it is particularly important to bear in mind that differences between mRNA and protein expression levels are possible with respect to intensities and to prognostic relevance. These differences underline the complementarity or synergy between expression measurements from cDNA arrays and IHC on TMA, and also the need for other high-throughput technologies such as cDNA arrays containing alternatively spliced transcripts,[39] protein arrays,[40] and *in situ* hybridizations on TMAs.[41] The combination of these complementary approaches will accelerate even more the identification of new diagnostic and prognostic markers as well as new therapeutic targets and will improve the management of breast cancer patients.

Acknowledgments

We thank J. Hassoun, G. Houvenaegel, D. Maraninchi, and C. Mawas for encouragement; and I. Bièche (Unité propre de recherche et d'enseignement supérieur JE2195, Paris, France), J. M. Durey (Institut Paoli-Calmettes, Marseille, France), O. P. Kallioniemi (National Human Genome Research Institute, Bethesda, MD), S. Lakhani (The Breakthough Toby Robins Breast Cancer Research Centre, London, UK), R. Lidereau (E0017 INSERM, St-Cloud, France), and B. Puig (Institut Paoli-Calmettes, Marseille, France) for help and advice.

References

1. Bertucci F, Houlgatte R, Benziane A, Granjeaud S, Adelaide J, Tagett R, Loriod B, Jacquemier J, Viens P, Jordan B, Birnbaum D, Nguyen C: Expression profiling in primary breast carcinomas using arrays of candidate genes. Hum Mol Genet 2000, 9:2981–2991
2. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pientas KJ, Rubin MA, Chinnaiyan AM: Delineation of prognostic biomarkers in prostate cancer. Nature 2001, 412:822–825
3. Shipp MA, Ross KN, Tamayo P, Weng A, Kutok JL, Aguiar RCT, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, Ray TS, Koval MA, Last KW, Norton A, Lister TA, Mesirov J, Neuberg DS, Lander ES, Aster JC, Golub TR: Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 2002, 8:68–74
4. Van't Veer LJ, Dai H, van de Vijver M, He YD, Hart AAM, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002, 415:530–535
5. Bertucci F, Houlgatte R, Nguyen C, Viens P, Jordan B, Birnbaum D: Gene expression profiling of cancer using DNA arrays: how far from the clinic? Lancet Oncol 2001, 2:674–682
6. Lakhani S, Ashworth A: Microarray and histopathological analysis of tumours: the future and the past? Nat Rev Cancer 2001, 1:151–157
7. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: Tissue microarrays for high-throughput molecular profiling of tumors specimens. Nat Med 1998, 4:844–847
8. Moch H, Schraml P, Bubendorf L, Mirlacher M, Kononen J, Gasser T, Mihatsch MJ, Kallioniemi OP, Sauter G: High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. Am J Pathol 1999, 154:981–986
9. Richter J, Wagner U, Kononen J, Fijan A, Bruderer J, Schmid U, Ackerman D, Maurer R, Alund G, Knönagel H, Rist M, Wilber K, Anabitarte M, Hering F, Hardmeier T, Schönenberger A, Flury R, Jäger P, Fehr JL, Schrami P, Moch H, Mihatsch MJ, Gasser T, Kallioniemi OP, Sauter G: High-throughput tissue microarray analysis of cyclin E gene amplification and overexpression in urinary bladder cancer. Am J Pathol 2000, 157:787–794
10. Bertucci F, Nasser V, Granjeaud S, Eisinger F, Tagett R, Adélaïde J, Loriod B, Benziane A, Giaconia A, Devilard E, Jacquemier J, Viens P, Nguyen C, Birnbaum D, Houlgatte R: Gene expression profiles of poor prognosis primary breast cancer correlate with survival. Hum Mol Genet 2002, 11:863–872
11. Möbus VJ, Moll R, Gerharz CD, Kieback DG, Merk O, Runnebaum IB, Linner S, Dreher L, Grill HJ, Kreienberg R: Differential characteristics of two new tumorigenic cell lines of human breast carcinoma origin. Int J Cancer 1998, 77:415–423
12. Theillet C, Adélaïde J, Louason G, Bonnet-Dorion F, Jacquemier J, Adnane J, Longy M, Katsaros D, Sismondi P, Gaudray P, Birnbaum D: FGFR1 and PLAT genes and DNA amplification at 8p12 in breast and ovarian cancers. Genes Chromosom Cancer 1993, 7:219–226
13. Bertucci F, Van Hulst S, Bernard K, Loriod B, Granjeaud S, Tagett R, Starkey M, Nguyen C, Jordan B, Birnbaum D: Expression scanning of an array of growth control genes in human tumor cell lines. Oncogene 1999, 18:3905–3912
14. Bertucci F, Bernard K, Loriod B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K, Jordan BR: Sensitivity issues in DNA array-based expression measurements: advantages of Nylon microarrays for small samples. Hum Mol Genet 1999, 8:1715–1722
15. Bièche I, Franc B, Vidaud D, Vidaud M, Lidereau R: Analyses of MYC, ERBB2 and CCND1 genes in benign and malignant thyroid follicular cell tumors by real-time polymerase chain reaction. Thyroid 2001, 11:147–152
16. Mitas M, Mikhitarian K, Walters C, Baron PL, Elliott BM, Brothers TE, Robison JG, Metcalf JS, Palesch YY, Zhang Z, Gillanders WE, Cole DJ: Quantitative real-time RT-PCR detection of breast cancer micrometastasis using a multigene marker panel. Int J Cancer 2001, 93: 162–171
17. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, Van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, Borresen-Dale AL: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001, 98:10869–10874
18. Ahr A, Karn T, Solbach C, Seiter T, Strebhardt K, Holtrich U, Kaufmann M: Identification of high risk breast-cancer patients by gene expression profiling. Lancet 2002, 359:131–132
19. Bertucci F, Eisinger F, Houlgatte R, Viens P, Birnbaum D: Gene expression profiling of breast cancer and prognosis. Lancet 2002, 360:173–174
20. Conte N, Charafe-Jauffret E, Delaval B, Adélaïde J, Ginestier C, Geneix J, Isnardon D, Jacquemier J, Birnbaum D: Carcinogenesis and translational controls: TACC1 is down-regulated in human cancers and associates with mRNA regulators. Oncogene 2002, 21: 5619–5630
21. Ugolini F, Adélaïde J, Charafe-Jauffret E, Nguyen C, Jacquemier J, Jordan B, Birnbaum D, Pébusque MJ: Differential expression assay of chromosome arm 8p genes identifies Frizzled-related (FRP1/FRZB) and fibroblast growth factor receptor 1 (FGFR1) as candidate breast cancer genes. Oncogene 1999, 18:1903–1910
22. Rahn JJ, Dabbagh L, Pasdar M, Hugh JC: The importance of MUC1 cellular localization in patients with breast carcinoma. Cancer 2001, 91:1973–1982
23. Hoos A, Cordon-Cardo C: Tissue microarray profiling of cancer specimens and cell lines: opportunities and limitations. Lab Invest 2001, 81:1331–1338
24. Nocito A, Kononen J, Kallioniemi OP, Sauter G: Tissue microarrays (TMAs) for high-throughput molecular pathology research. Int J Cancer 2001, 94:1–5
25. Rimm DL, Camp RL, Charette LA, Olsen DA, Provost E: Amplification

of tissue by construction of tissue microarrays. Exp Mol Pathol 2001, 70:255–264
26. Camp RL, Charette LA, Rimm DL: Validation of tissue microarray technology in breast carcinoma. Lab Invest 2000, 80:1943–1949
27. Hoos A, Urist MJ, Stojadinovic A, Mastorides S, Dudas ME, Leung DHY, Kuo D, Brennan MF, Lewis JL, Cordon-Cardo C: Validation of the tissue microarray for immunohistochemical profiling of cancer specimens using the example of human fibroblastic tumors. Am J Pathol 2001, 158:1245–1251
28. Torhorst J, Bucher C, Kononen J, Haas P, Zuber M, Kochli OR, Mross F, Dieterich H, Moch H, Mihatsch M, Kallioniemi OP, Sauter G: Tissue microarray for rapid linking of molecular changes to clinical endpoints. Am J Pathol 2001, 159:2249–2256
29. Rubin MA, Dunn R, Strawderman M, Pienta KJ: Tissue microarray sampling strategy for prostate cancer biomarker analysis. Am J Surg Pathol 2002, 26:312–319
30. Rubin MA, Zhou M, Dhanasekaran SM, Varambally S, Barrette TR, Sanda MG, Pienta KJ, Ghosh D, Chinnaiyan AM: Alpha-methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. JAMA 2002, 287:1662–1670
31. Matthias LJ, Gotis-Graham I, Underwood PA, McNeil HP, Hogg PJ: Identification of monoclonal antibodies that recognize different disulfide bonded forms of thrombospondin 1. Biochem Biophys Acta 1996, 1296:138–144
32. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci USA 1999, 96:9212–9217
33. Mousses S, Kallioniemi A, Kauraniemi P, Elkahloun A, Kallioniemi OP: Clinical and functional target validation using tissue and cell microarrays. Curr Opin Chem Biol 2001, 6:97–101
34. Bubendorf L, Kolmer M, Kononen J, Koivisto P, Mousses S, Chen Y, Mahlamaki E, Schraml P, Moch H, Willi N, Elkahloun AG, Pretlow TG, Gasser TC, Mihatsch MJ, Sauter G, Kallioniemi OP: Hormone therapy failure in human prostate cancer: analysis by complementary DNA and tissue microarrays. J Natl Cancer Inst 1999, 91:1758–1764
35. Sallinen SL, Sallinen PK, Haapasalo HK, Helin HJ, Helen PT, Schraml P, Kallioniemi OP, Kononen J: identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques. Cancer Res 2000, 60:6617–6622
36. Su AI, Welsh JB, Sapinoso LM, Kern SG, Dimitrov P, Lapp H, Schultz PG, Powell SM, Moskaluk CA, Frierson HF, Hampton GM: Molecular classification of human carcinomas by use of gene expression signatures. Cancer Res 2001, 61:7388–7393
37. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, Wilfond B, Borg A, Trent J: Gene-expression profiles in hereditary breast cancer. N Engl J Med 2001, 344:539–548
38. McGuckin MA, Walsh MD, Hohn BG, Ward BG, Wright RG: Prognostic significance of MUC1 epithelial mucin expression in breast cancer. Hum Pathol 1995, 26:432–439
39. Yeakley JM, Fan JB, Doucet D, Luo L, Wickham E, Ye Z, Chee MS, Fu XD: Profiling alternative splicing on fiber-optic arrays. Nat Biotechnol 2002, 20:353–358
40. Emili AQ, Cagney G: Large-scale functional analysis using peptide or protein arrays. Nat Biotechnol 2000, 18:393–397
41. Fejzo MS, Slamon DJ: Frozen tumor tissue microarray technology for analysis of tumor RNA, DNA, and proteins. Am J Pathol 2001, 159: 1645–1650

1: Int J Mol Med. 1998 May;1(5):855-61.

p185 overexpression in 220 samples of breast cancer undergoing primary surgery: comparison with c-erbB-2 gene amplification.

Dalifard I, Daver A, Goussard J, Lorimier G, Gosse-Brun S, Lortholary A, Larra F.

Laboratoire de Radioanalyse, Centre Paul Papin, 49033 Angers Cedex 01, France.

In breast cancer, DNA amplification of the oncogene c-erbB-2, encoding for the p185 protein, is associated with a poor prognosis. A retrospective study o a population of 220 cases of primary breast cancer permitted a quantitative measure of p185 oncoprotein overexpression by an immunoenzymetric assay and the determination of c-erbB-2 amplification by the Southern blot method A correlation existed between the two measurements (r=0.85) using the double cut-off: DNA 2 copies and p185 400 U/mg protein, and only 2.7% of the cases were discordant. 13.2% of the tumors showed p185 overexpression The percentage of tumors overexpressing p185 was significantly different between the groups with amplified and non-amplified c-erbB-2. We observec a significant correlation between p185 levels and tumor grade (p=0.03), and an inverse correlation with hormonal receptors (p=0.0001). The p185 assay could be an additional prognostic factor to better define patient subgroups with node negative, grade II, and positive or negative hormonal receptors.

PMID: 9852307 [PubMed - indexed for MEDLINE]

Biochem. 189, 475–486 (1990)
FEBS 1990.

A transcribed gene, containing a variable number of tandem repeats, codes for a human epithelial tumor antigen

DNA cloning, expression of the transfected gene and over-expression in breast cancer tissue

Mara HAREUVENI[1, 2], Ilan TSARFATY[1], Joseph ZARETSKY[1], Phillip KOTKES[1], Judith HOREV[1], Sheila ZRIHAN[1], Mordechai WEISS[1], Stephen GREEN[2], Richard LATHE[2], Iafa KEYDAR[1] and Daniel H. WRESCHNER[1]

[1] Department of Microbiology, Faculty of Life Sciences, Tel Aviv University, Israel
[2] Laboratoire de Genetique Moleculaire des Eucaryotes du Center National Récherche Scientifique, Institut de Chimie Biologique, Faculte de Medecine, Strasbourg, France

(Received July 28/December 13, 1989) – EJB 89 0945

A monoclonal antibody, H23, that specifically recognizes a breast-tumor-associated antigen, was used to isolate a cDNA insert that codes for the antigenic epitope. Nucleotide sequencing of this cDNA, as well as a longer 850-bp cDNA insert, shows that they are composed of 60-bp (G + C)-rich tandem repeating units. The coding strand was determined and codes for a proline-rich 20-amino-acid repeat motif. A comparison of the highly conserved repeat unit with the deduced flanking amino acid sequences demonstrates conservation of specific subregions of the repeat consensus within the flanking amino acids. Hybridization of the 60-bp cDNA probe with RNAs extracted from a variety of primary and metastatic human tumors yields relatively high levels of hybrid with the breast carcinomas, as compared to lower hybrid levels with RNAs from other epithelial tumors. RNA extracted from breast tissue adjacent to the tumor or from benign breast tumors, demonstrates low or undetectable levels of hybridization. Probing Southern blots with the 60-bp repeat shows that the tumor antigen is highly polymorphic and contains a variable number of tandem repeats (VNTRs). The VNTR nature of the gene was confirmed by probing Southern blots with unique genomic sequences that are physically linked to an isolated gene fragment that also contains the tandem repeat array. Mouse cells transfected with this gene fragment produce tumor antigen that is readily detected by H23 monoclonal antibodies. The allelic forms seen in 10 different primary human tumors demonstrate 100% concordance with the various mRNA species expressed. These studies are extended to the protein forms detected by immunoblot analyses that show both a correlation of the expressed tumor antigen species with the allelic forms as well as significantly increased expression in breast cancer tissue. The above studies unequivocally establish the over-expression of a VNTR gene coding for an epithelial tumor antigen in human breast cancer tissue.

The isolation and characterization of proteins that are aberrantly expressed in human tumor tissues may elucidate cellular mechanisms leading to malignancy and also be of significant clinical importance. To identify breast-tumor-associated markers, we have established a human breast cancer cell line, T47D, that has been extensively studied and retains many characteristics of primary human breast tumors [1]. Monoclonal antibodies (mAb) were prepared against particulate antigens released by these cells and screened for their specificity by the immunohistochemical staining of breast tissue sections. One mAb, designated H23, stained cytoplasmically 91% of all malignant breast tumors analyzed, whereas little or no cytoplasmic staining was observed in normal and benign breast tissues [2].

Correspondence to D. H. Wreschner, Department of Microbiology, Tel Aviv University, Ramat Aviv, I-69978, Israel

Abbreviations. mAb, monoclonal antibody; H23 Ag, epithelial tumor antigen recognized by H23 monoclonal antibodies; ORF, open reading frame; VNTR, variable number of tandem repeats.

Note. The novel nucleotide sequence data published here and in the preceding paper in this journal have been deposited with the EMBL sequence data bank and are available under the accession number X52228 and X52229. The novel amino acid sequence data have been deposited with the EMBL sequence data bank.

Other groups have also described mAbs reactive with high-molecular-mass glycoproteins that are aberrantly expressed in epithelial tumors and especially in breast cancer [3–12]. Several of these mAbs, namely DF3, HMFG-1, HMFG-2 and SM-3, were used to isolate cDNA clones that express the immunereactive epitope [13–15]. The cDNAs isolated all contain tandem 60-bp repeat units [13–15] that code for a 20-amino-acid repeat motif rich in proline, serine, threonine and alanine [14]. Southern blots probed with the 60-bp repeat cDNA insert show that the gene is highly polymorphic and correlates with the polymorphism observed in the protein products [13, 16]. These results suggest the codominant autosomal expression of a gene that contains a variable number of tandem repeats (VNTR).

Besides the 60-bp repeat unit, little has been known regarding the unique non-repeat sequences of the tumor antigen cDNA and its gene. Indeed, the aforementioned studies on the molecular structure of the epithelial antigen, including all Southern and Northern blot analyses, have been performed solely with the 60-bp cDNA repeat unit [13–16]. To study the breast-tumor-associated antigen recognized by H23 mAbs (H23 Ag) on a molecular level, a gt11 cDNA expression library prepared from T47D mRNA was screened with these

mAbs. A cDNA insert that codes for the immunereactive epitope was isolated and sequenced. This insert, as well as a longer 850-bp cDNA insert, are composed of 60-bp tandem repeats, similar to those previously reported [13–16]. We have recently increased our knowledge beyond the confines of the 60-bp repeat units by isolating almost full-length cDNAs that contain unique non-repeat sequences located 5′ and 3′ to the tandem repeat array and code for the complete epithelial tumor antigen [16a].

We have extended these studies and report here the determination of the VNTR nature of the gene by analyzing Northern and Southern blots with probes consisting not only of the 60-bp cDNA repeat [13–16] but also with probes derived from unique non-repeat genomic sequences. These investigations were performed on nucleic acids isolated from primary human tissues and are therefore relevant to the *in-vivo* situation. In addition, we demonstrate the over-expression of mRNA coding for the tumor antigen and of the antigen itself in primary breast cancer tissues. The coding strand of the 60-bp repeat unit has been determined and a comparative analysis of the tumor antigen unique non-repeat amino acid sequences with the 20-amino-acid repeat motif is presented. Finally, by transfecting cells with the isolated gene coding for the tumor antigen, stable mouse cell transfectants have been established that express the human-breast-tumor-associated antigen.

The findings reported here unequivocally establish the over-expression in human breast cancer tissue of a VNTR gene that codes for an epithelial tumor antigen.

MATERIALS AND METHODS

Plating of the recombinant cDNA phage and library screening with H23 mAb

The randomly primed λgt11 cDNA expression library [17] was prepared from poly(A)-rich RNA of T47D cells, a human breast carcinoma cell line [2]. Approximately 10^6 phages were plated on *Escherichia coli* strain Y1090 and the resulting plaques were screened for expression of crossreacting galactosidase fusion protein, with 25 μg/ml H23 IgG as described elsewhere [18]. For the final detection of positive plaques, ^{125}I-protein A was used at a final concentration of 4×10^5 cpm/ml. Positive plaques were picked and rescreened repeatedly until all plaques were immunopositive. Most of the λ cDNA clones contained an insert of similar size and the clone with the longest insert, designated 3b, was thus obtained and used for Northern hybridization assays.

DNA hybridization of cDNA library

The cDNA library replica-plated on nylon membranes (Amersham, England) was probed with cDNA inserts labelled by nick translation [19] to a specific activity $2-5 \times 10^8$ cpm/μg and a final concentration of $1-2 \times 10^6$ Cerenkov cpm/ml. The replica blots were prehybridized and probed at 42°C for 15 h in 50% formamide, 5 × NaCl/Cit (1 × NaCl/Cit is 150 mM NaCl, 15 mM sodium citrate, pH 7.0), 0.1% polyvinylpyrrolidone, 0.1% Ficoll, 0.2% SDS and 100 μg/ml denatured salmon sperm DNA.

Following hybridization, the blots were washed at 65°C for 2–4 h with several changes of 2 × NaCl/Cit, 0.2% SDS following by stringent washing at 65°C (2 × 30 min) with 0.2 × NaCl/Cit, 0.5% SDS. The washed blots were exposed to Agfa Gevaert X-ray films at −70°C using curix-special intensifying screens.

Southern blot DNA analysis

High-molecular-mass DNA was isolated from powdered surgically removed frozen (−70°C) tissues by incubating overnight at 50°C in 200 μg/ml proteinase K, 100 mM NaCl, 10 mM Tris/HCl pH 7.5, 1 mM EDTA followed by phenol/chloroform and one chloroform extraction. The DNA was spooled onto glass rods after the addition of 0.2 M NaCl (final concentration) and 1 vol. absolute ethanol at −20°C. The spooled DNA was rinsed with 70% ethanol, briefly dried, resuspended in double distilled water and kept at −20°C. The DNA (100 μg/ml final concn) was incubated with the appropriate restriction enzymes (approximately 5 units enzyme/μg DNA) overnight at 37°C followed by ethanol precipitation at −20°C. When double digestions were performed, DNA was incubated with one enzyme followed by ethanol precipitation, resuspension and then digestion with the second enzyme. 10–20 μg restricted DNA was electrophoresed on 0.8% agarose gels in recirculating Tris/acetate/EDTA buffer, followed by staining with ethidium bromide and washing in 1.5 M NaCl, 0.5 M NaOH for 30 min. Southern transfer to nylon membranes (Amersham, England) was performed in 1.5 M NaCl, 0.25 M NaOH. The blot was irradiated with ultraviolet light for 3 min, followed by baking at 80°C for 2 h. Prehybridization, hybridization and washing were as described above.

RNA analysis

RNA was extracted from surgically removed frozen tissues using the guanidinium thiocyanate/cesium chloride method [20]. Poly(A)-rich RNA was purified by oligo(dT)-cellulose chromatography [21]. For dot-blot analysis, 15 μg of each sample of total RNA was applied with gentle vacuum in 200 μl of 2 × NaCl/Cit to a Gelman nylon membrane using the BRL dot-blot apparatus. The RNA samples were covalently attached to the nylon membrane by ultraviolet irradiation followed by baking at 80°C under vacuum.

For Northern analysis 40 μg of each total RNA sample or 4 μg of poly(A)-rich selected RNA were subjected to electrophoresis on a 1.4% agarose gel under glyoxal/dimethylsulfoxide-denaturing conditions using Tris/acetate/EDTA as the running buffer. Subsequent to 50 mM NaOH treatment and washings in 2 × NaCl/Cit, the gels were stained by ethidium bromide and Northern blotted to Gelman nylon membranes [21].

Northern and RNA dot-blot hybridizations

The blots obtained as previously described, were prehybridized and probed at 42°C for 16 h in 50% formamide, 5 × NaCl/Cit, 0.1% polyvinylpyrrolidone, 0.1% Ficoll, 0.2% SDS and 100 μg/ml denatured salmon sperm DNA with cDNA inserts labelled by nick translation [19] to a specific activity of $2-5 \times 10^8$ cpm/μg. A final concentration of $1-2 \times 10^6$ Cerenkov cpm/ml was used. Following hybridization, the blots were washed at 65°C for 2–4 h with several changes of 2 × NaCl/Cit, 0.2% SDS followed by stringent washing at 65°C (2 × 30 min) with 0.2 × NaCl/Cit, 0.5% SDS. Quantification of the hybridization intensity was performed with the LKB 2222-020 Ultrascan XL II laser densitometer. Bound probes were removed by washing blots in hybridization buffer

65°C for 60 min and the membr[illegible]re then rehybridized with a different probe under similar conditions.

Construction of eukaryotic expression vector coding for H23 Ag

The *Xmn*I – *Eco*RI genomic fragment (see Fig. 6) was inserted into the eukaryotic expression vector pCL642 (this vector will be described in detail in a separate publication). Briefly, pCL642 is composed of the promoter region (1.4 kb) isolated from the mouse housekeeping gene coding for 3-hydroxy-3-methylglutaryl-coenzyme-A reductase. The promoter is followed by the untranslated first exon and intron (0.7 kb and 3.5 kb) derived also from this reductase gene. The *Xmn*I site of the *Xmn*I – *Eco*RI genomic fragment coding for the tumor antigen was blunt-end-ligated to an *Eco*RV site located in a polylinker immediately downstream to the reductase intron. The *Eco*RI site was ligated to the *Kpn*I site of the polylinker via an *Eco*RI – *Kpn*I adaptor. A 123-bp fragment containing the SV40 poly(A) signal sequence is situated immediately 3′ to the polylinker. The construct pCL642/*Xmn*I – *Eco*RI (10 μg) was contransfected with 1 μg pAG60(G418R) plasmid [22] into either MM5tC3H cells (from American Tissue Culture Collection) or FR3T3 ras-1 cells [23] using a modification [24] of the calcium phosphate precipitation method [25]. Cells were selected for G418 resistance (Geniticin 500 μg/ml) and loci were picked and subcultured.

For the detection of tumor antigen, the transfected cells were grown on coverslips and immunohistochemically stained with H23 mAbs. Control cells were either transfected with the pAG60 plasmid alone or with an irrelevant gene.

Nucleotide sequencing

Sequencing was accomplished using the dideoxynucleotide chain-termination method [26]. Restriction fragments of the cDNA inserts were subcloned into M13 and both strands were sequenced. The ssDNA was primed with either the M13 universal primer or synthetic oligonucleotides prepared according to known sequences. The analysis of the sequence was performed using the Beckman MicroGenie program.

Radioactive labelling of DNA probes

Double-stranded DNA probes were radioactively labelled with [α-^{32}P]dCTP either by nick translation or random oligonucleotide multipriming using commercially available kits (BRL, USA, and Amersham, England, respectively). All DNA probes used here were purified inserts that were isolated by agarose gel electrophoresis. Single-stranded oligonucleotides were 5′-end labelled by incubating with [γ-^{32}P]ATP and polynucleotide kinase. All labelled probes were purified from non-incorporated nucleotide by passage through Sephadex G-100 columns.

Oligonucleotide synthesis

Oligonucleotides were prepared at the Macromolecule Synthesis Service Unit (Department of Organic Chemistry, Weizman Institute of Science) by Dr Ora Goldberg using an Automated Applied Biosystems synthesizer. Following synthesis, the oligonucleotides were electrophoretically purified on acrylamide/urea gels.

Immunoblotting

Protein samples denatured by boiling in SDS buffer containing mercaptoethanol were analyzed on 3–15% linear gradient SDS/acrylamide gels as previously described [27]. Following electrophoresis, the gel was equilibrated in transfer buffer (Tris/glycine) and electrotransferred for 3 h at 1 A to nitrocellulose filters. The filters were blocked in NaCl/P_i (150 mM NaCl, 15 mM sodium phosphate, pH 7.0) containing 5% skimmed milk (Blotto) followed by incubation with antibody in Blotto. The filters were washed in NaCl/P_i and then reacted with ^{125}I-protein A (Amersham, England) dissolved in Blotto.

Monoclonal antibodies

Monoclonal antibodies (mAbs) were prepared against particulate antigens released into the medium by T47D breast carcinoma cells, using established procedures. The monoclonal antibodies obtained were screened against paraffin-embedded sections of benign and malignant breast tissue with the immunoperoxidase-staining technique and one of the mAbs designated H23 [2] was selected and used in this study.

RESULTS

Sequence of cDNA coding for epitope recognized by H23 monoclonal antibodies

The gt11 cDNA expression libraries prepared with mRNA isolated from either T47D or MCF7 [17], both human breast carcinoma cell lines, were probed with the monoclonal antibody H23. Libraries obtained by priming poly(A)-rich RNA with oligo(dT), as well as with random nucleotide oligomers, were investigated. Recombinant clones immunoreactive with H23, were obtained at a frequency of approximately 1 in 2000 in the amplified libraries and the cDNA inserts of all clones analyzed revealed a size of approximately 220–240 bp. Both random oligomer-primed as well as oligo(dT)-primed libraries from the MCF7 and T47D cell lines revealed similar-sized inserts.

Nucleotide sequencing of one such representative cDNA insert, termed 3b, indicates that it is (G + C)-rich with strand preference for the G or C nucleotides. Inspection of the 225-bp sequence shows that it is composed of a 60-nucleotide tandem repeat unit which is remarkably conserved with only very few substitutions occurring between the different units (Fig. 1A).

Longer cDNA inserts contain the same 60-bp tandem repeat unit

In order to obtain longer cDNA clones, the 3b cDNA insert was used to reprobe the library by DNA/DNA hybridization. Several recombinant phages with longer inserts were obtained, the longest of which is approximately 850 bp. Nucleotide sequencing of this insert indicated that it is solely composed of the same tandem 60-nucleotide repeat unit. Similarly other longer cDNA inserts obtained by 3b cDNA probing of the library are also only composed of the tandem repeating unit.

Restriction enzyme digestion of the isolated 850-bp insert with *Sma*I (CCCGGG) completely reduces it to 60-bp fragments (data not shown) thus indicating that *Sma*I sites appear at 60-bp intervals.

A comparison of the repeat units found in the 3b cDNA with those present in the 850-bp insert (Fig. 1A, B and C)

°C for 60 min and the membran[illegible] then rehybridized with a different probe under similar c[illegible]itions.

Construction of eukaryotic expression vector coding for H23 Ag

The *Xmn*I – *Eco*RI genomic fragment (see Fig. 6) was inserted into the eukaryotic expression vector pCL642 (this vector will be described in detail in a separate publication). Briefly, pCL642 is composed of the promoter region (1.4 kb) isolated from the mouse housekeeping gene coding for 3-hydroxy-3-methylglutaryl-coenzyme-A reductase. The promoter is followed by the untranslated first exon and intron (0.7 kb and 3.5 kb) derived also from this reductase gene. The *Xmn*I site of the *Xmn*I – *Eco*RI genomic fragment coding for the tumor antigen was blunt-end-ligated to an *Eco*RV site located in a polylinker immediately downstream to the reductase intron. The *Eco*RI site was ligated to the *Kpn*I site of the polylinker via an *Eco*RI – *Kpn*I adaptor. A 123-bp fragment containing the SV40 poly(A) signal sequence is situated immediately 3′ to the polylinker. The construct pCL642/*Xmn*I – *Eco*RI (10 μg) was contransfected with 1 μg pAG60(G418R) plasmid [22] into either MM5tC3H cells (from American Tissue Culture Collection) or FR3T3 ras-1 cells [23] using a modification [24] of the calcium phosphate precipitation method [25]. Cells were selected for G418 resistance (Geniticin 500 μg/ml) and loci were picked and subcultured.

For the detection of tumor antigen, the transfected cells were grown on coverslips and immunohistochemically stained with H23 mAbs. Control cells were either transfected with the pAG60 plasmid alone or with an irrelevant gene.

Nucleotide sequencing

Sequencing was accomplished using the dideoxynucleotide chain-termination method [26]. Restriction fragments of the cDNA inserts were subcloned into M13 and both strands were sequenced. The ssDNA was primed with either the M13 universal primer or synthetic oligonucleotides prepared according to known sequences. The analysis of the sequence was performed using the Beckman MicroGenie program.

Radioactive labelling of DNA probes

Double-stranded DNA probes were radioactively labelled with [α-^{32}P]dCTP either by nick translation or random oligonucleotide multipriming using commercially available kits (BRL, USA, and Amersham, England, respectively). All DNA probes used here were purified inserts that were isolated by agarose gel electrophoresis. Single-stranded oligonucleotides were 5′-end labelled by incubating with [γ-^{32}P]ATP and polynucleotide kinase. All labelled probes were purified from non-incorporated nucleotide by passage through Sephadex G-100 columns.

Oligonucleotide synthesis

Oligonucleotides were prepared at the Macromolecule Synthesis Service Unit (Department of Organic Chemistry, Weizman Institute of Science) by Dr Ora Goldberg using an Automated Applied Biosystems synthesizer. Following synthesis, the oligonucleotides were electrophoretically purified on acrylamide/urea gels.

Immunoblotting

Protein samples denatured by boiling in SDS buffer containing mercaptoethanol were analyzed on 3–15% linear gradient SDS/acrylamide gels as previously described [27]. Following electrophoresis, the gel was equilibrated in transfer buffer (Tris/glycine) and electrotransferred for 3 h at 1 A to nitrocellulose filters. The filters were blocked in NaCl/P_i (150 mM NaCl, 15 mM sodium phosphate, pH 7.0) containing 5% skimmed milk (Blotto) followed by incubation with antibody in Blotto. The filters were washed in NaCl/P_i and then reacted with ^{125}I-protein A (Amersham, England) dissolved in Blotto.

Monoclonal antibodies

Monoclonal antibodies (mAbs) were prepared against particulate antigens released into the medium by T47D breast carcinoma cells, using established procedures. The monoclonal antibodies obtained were screened against paraffin-embedded sections of benign and malignant breast tissue with the immunoperoxidase-staining technique and one of the mAbs designated H23 [2] was selected and used in this study.

RESULTS

Sequence of cDNA coding for epitope recognized by H23 monoclonal antibodies

The gt11 cDNA expression libraries prepared with mRNA isolated from either T47D or MCF7 [17], both human breast carcinoma cell lines, were probed with the monoclonal antibody H23. Libraries obtained by priming poly(A)-rich RNA with oligo(dT), as well as with random nucleotide oligomers, were investigated. Recombinant clones immunoreactive with H23, were obtained at a frequency of approximately 1 in 2000 in the amplified libraries and the cDNA inserts of all clones analyzed revealed a size of approximately 220–240 bp. Both random oligomer-primed as well as oligo(dT)-primed libraries from the MCF7 and T47D cell lines revealed similar-sized inserts.

Nucleotide sequencing of one such representative cDNA insert, termed 3b, indicates that it is (G + C)-rich with strand preference for the G or C nucleotides. Inspection of the 225-bp sequence shows that it is composed of a 60-nucleotide tandem repeat unit which is remarkably conserved with only very few substitutions occurring between the different units (Fig. 1A).

Longer cDNA inserts contain the same 60-bp tandem repeat unit

In order to obtain longer cDNA clones, the 3b cDNA insert was used to reprobe the library by DNA/DNA hybridization. Several recombinant phages with longer inserts were obtained, the longest of which is approximately 850 bp. Nucleotide sequencing of this insert indicated that it is solely composed of the same tandem 60-nucleotide repeat unit. Similarly other longer cDNA inserts obtained by 3b cDNA probing of the library are also only composed of the tandem repeating unit.

Restriction enzyme digestion of the isolated 850-bp insert with *Sma*I (CCCGGG) completely reduces it to 60-bp fragments (data not shown) thus indicating that *Sma*I sites appear at 60-bp intervals.

A comparison of the repeat units found in the 3b cDNA with those present in the 850-bp insert (Fig. 1A, B and C)

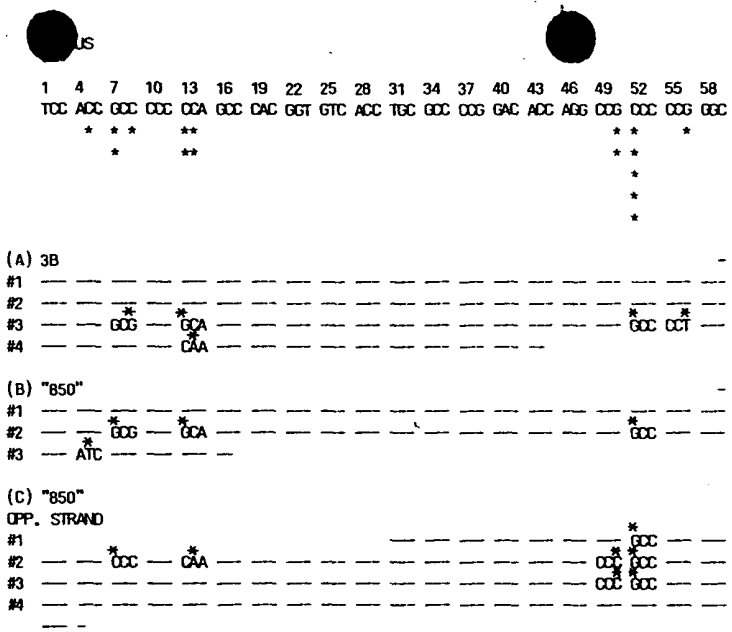


Fig. 1. *Nucleotide sequence of cDNAs that code for the epitope recognized by H23 mAbs and contain 60-bp tandem repeat units.* The gt11 cDNA expression library was screened with H23 mAbs as described in Methods and the cDNA insert (indicated as 3B but referred to in the text as 3b) obtained from a positive purified recombinant phage was subcloned in M13 vectors in both orientations and sequenced (A). The 3b cDNA insert was purified, nick-translated and used to reprobe the library under stringent hybridization conditions as described in Methods. The longest cDNA insert ('850', i.e. 850 bp) thus obtained was subcloned in M13 and both strands were partially sequenced (B and C). Only the C-rich strand is presented. The consensus sequence of the 60-bp repeat unit is shown at the top of the figure. Nucleotides in the repeat units identical to this sequence are indicated with dashes whilst substitutions are shown by an asterisk

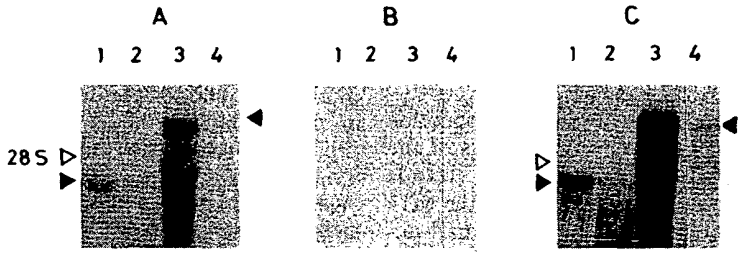


Fig. 2. *Northern blot analysis of human breast tumor RNA samples with 3b cDNA probe and synthetic complementary oligonucleotides derived from the repeating unit.* RNA was extracted from human breast tumor tissue (lanes 1 and 3) or adjacent 'normal' breast tissue (lanes 2 and 4) from two separate individuals (lanes 1 + 2 and 3 + 4) and analyzed by glyoxal agarose gel electrophoresis followed by Northern blotting to nylon membranes and hybridization with (A) the 3b cDNA probe, (B) the C-rich oligonucleotide 5′ AGCCCACGGTGTCACCTCGGCCCCGGACA 3′ identical to nucleotides 15–43 of the concensus sequence presented in Fig. 1 A and (C) the complementary G-rich oligonucleotide 5′ TGTCCGGGGCCGAGGTGACACCGTGGGCT 3′. The probe used in A was radioactively labelled by nick translation whilst those used in B and C were end-labelled by polynucleotide kinase and [γ-^{32}P]ATP as described in Methods. The blots were stringently washed and autoradiographed at −70°C. The full arrow to the left of the figure indicates the 3.6-kb hybridizing mRNA species whilst that on the right points to the 6.0-kb mRNA species detected in the other sample. The open arrow indicates the position of 28S rRNA

demonstrates remarkable conservation with very few nucleotide substitutions occurring between the repeats.

Coding strand of the tandem repeat unit

To determine the coding strand of the tandem repeat unit, Northern blots were probed with synthetic oligonucleotides complementary to either strand of the repeat unit. Probing a Northern blot containing RNA isolated from human breast samples (both tumor and adjacent 'normal' tissues) shows that the G-rich synthetic oligonucleotide hybridizes to mRNA species (Fig. 2). An identical hybridization pattern was observed when 3b cDNA was used to probe the same blot (Fig. 2). The breast tumor tissue has in the one case a hybridizing mRNA species of 6.5 kb whilst the second sample shows a single band at 3.6 kb. The corresponding RNA samples from adjacent 'normal' tissue are identically sized but much reduced in amount. In contrast to the above results, no hybridization at all is seen with the second complementary C-rich oligonucleotide (Fig. 2). These findings confirm that RNA species containing multiple 60-nucleotide tandem repeats are *bonafide* transcripts. Moreover the orientation of transcription is demonstrated and the C-rich strand of the cDNA insert is the coding strand.

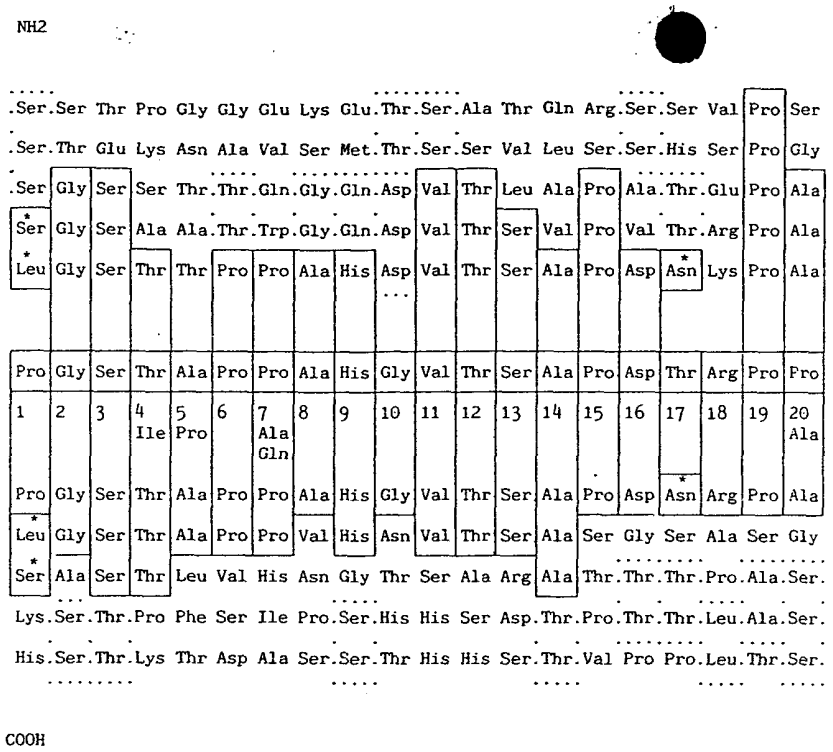


Fig. 3. *Comparative analysis of the flanking amino acid sequences with the 20-amino-acid repeat motif.* The amino acid sequence of the repeat motif is presented in the central boxed region and numbered from 1 to 20. The alternative amino acids that occur due to variations in the consensus sequence are indicated below the numbers. The 100 amino acids flanking the repeat motif on the amino and carboxyl terminals are shown (NH_2 and COOH, respectively). Flanking amino acids that are identical with the repeat motif are boxed in by the full line, whereas the flanking amino acids that appear in the same position every 20 amino acids are boxed in by a series of dots. Amino acids that vary from the repeat motif and appear at the same positions on either side of the repeat motif are indicated by * and are boxed in

Comparative analysis of the flanking amino acid sequences with the 20-amino-acid repeat motif

The determined coding strand of the 60-bp cDNA could be translated in all three reading frames. As almost full-length cDNAs coding for the H23 Ag have recently been isolated [16a], the correct reading frame of the repeat motif could be readily identified (Fig. 3). The high level of nucleotide conservation amongst the various repeat units is reflected in the repeat-unit amino acid sequences (Fig. 3). The studies reported here (see also below) show that the tumor antigen has the unusual structure of highly conserved repeat units that compose at least 50% of the protein molecule. It is therefore of considerable interest to compare the similarity of flanking non-repeat amino acid sequences with the 20-amino-acid repeat motif itself.

Several possibilities may be envisaged. (a) An abrupt break may occur in the continuity of the repeat motif and no similarity exist between the flanking amino acid sequences and the repeat units. (b) Some degree of similarity may exist between the flanking amino acid sequences and the repeat motif that declines with increasing distance away from the repeat array. Or (c) the flanking amino acid sequences may retain similarity only with specific amino acids or regions of the consensus repeat motif.

The comparative analysis (Fig. 3) shows that indeed similarity exists between the flanking amino acid sequences and the repeat motif itself. However, this similarity is confined to specific subregions such as the Val-Thr-Ser and Gly-Ser peptides at residues 11 – 13 and 2 – 3, respectively, and occurs in the flanking amino acid sequences on both sides of the repeat motif. On the other hand, certain amino acid residues are conserved asymmetrically, i.e. either upstream or downstream to the repeat motif. Significant conservation in the amino-terminal flanking sequences occurs with the proline residues (15 and 19) and the alanine residue (20), whereas threonine (4) and alanine (14) are appreciably conserved only in the carboxyl-terminal flanking sequences. The conservation of the proline residue (19) in the upstream flanking sequences is particularly remarkable as it is located in the same position for 82 amino acids upstream to the repeat motif. Of further note are the amino acids that diverge from the repeat motif: the asparagine residue that replaces threonine (17) does so on both sides flanking the repeat motif. Furthermore, proline (1) is replaced by leucine and followed 20 amino acids later by serine; it is indeed striking that identical changes occur both in the upstream and downstream flanking sequences.

At the present time, the significance of repeat-motif amino acid conservation, as well as the identical amino acid changes occurring on both sides of the repeat, is not known. They may impose certain structural constraints on the protein molecule or/and be related to a function involving specific subregions of the repeat motif.

Expression of tumor antigen mRNA in primary human tissues

The *in vivo* system studies were extended and we investigated the presence of mRNA species hybridizing with the 3b cDNA probe in a variety of benign and malignant human tissues by RNA dot blotting and Northern blot analysis. The initial dot-blot screening demonstrates very significant levels of hybridizing mRNA species in total RNA prepared from a

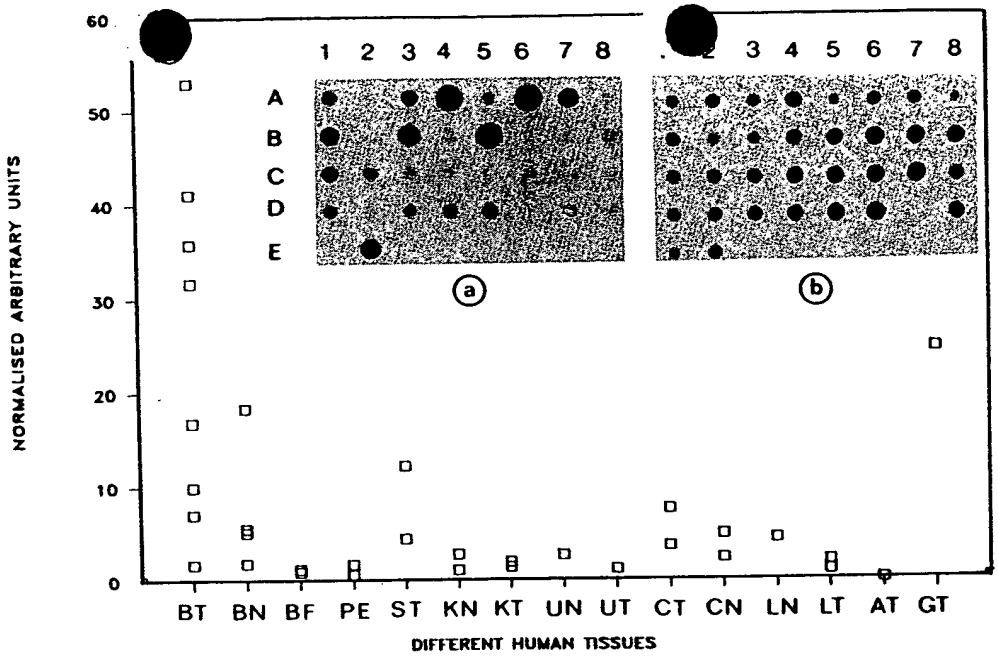


Fig. 4. *Levels of RNA species in human tissues hybridizing with the 3b cDNA probe.* Total RNA from different human tissues was dot-blotted and probed with (a) the 3b cDNA insert and (b) a cDNA insert (unpublished results) coding for part of human 18S ribosomal RNA. The key for the dot blots is as follows: (A1–8) BT2, BT4, BT5, BT6, BN7, BT7, BN9, BN10; (B1–8) BT10, BN12, BT12, BF13, BT15, BF16, PE1, PE2; (C1–8) ST1, ST2, KN1, KN2, KT2, UT1, UN1; (D1–8) CT1, CN1, CT2, LN1, LT1, blank, LT2; (E1–2) AT, GT1. Abbreviations used are for RNA extracted from: BT = breast adenocarcinoma, ST = gastric carcinoma, KT = hypernephroma, UT = transitional cell carcinoma, CT = colon adenocarcinoma, LT = lung tumor, AT = pheochromocytoma and GT = ovarian carcinoma. The corresponding 'N' samples (for example BN) represent RNA isolated from 'normal' tissue adjacent to the tumor. The same numbers indicate preparations from the same patient. The BF samples are breast fibroadenomas. PE samples are from pleural effusion metastatic cells of patients with advanced breast cancer. (c) The dot blots were scanned by laser densitometry using an LKB laser densitometer and the absorbance values obtained with the blot probed with the 3b cDNA probe were divided by the levels observed following 18S cDNA probing. This procedure resulted in a normalized arbitrary unit corresponding to each sample which is presented on the ordinate of the figure. Total RNA extraction, dot blotting, hybridization and washing conditions were as described in Materials and Methods. The blots were exposed to Agfa Curix X-ray films at −70°C with a Curix special intensifying screen

number of breast carcinomas (Fig. 4A). Several of these samples contain large quantities of mRNA capable of hybridizing with the 3b cDNA probe and a quantitative analysis demonstrates high levels of mRNA hybridization to the 3b cDNA probe (Fig. 4C). Significantly lower levels of hybridization are observed in RNA isolated from non-malignant breast tissue adjacent to the biopsied tumor sample. For example, Fig. 4A shows that RNA isolated from tissue adjacent to tumor in samples BN7 (B = breast, N = normal), BN10 and BN12 (dot-blot positions A5, A8 and B2, respectively) demonstrate hybridizing values of 5.5, 5.1 and 1.8 (normalized probe-specific hybridization) whereas 3b cDNA hybridization to RNA extracted from the corresponding tumor samples BT7 (B = breast, T = tumor), BT10 and BT12 (dot-blot positions A6, B1 and B3, respectively) shows considerably higher values of 53.0, 16.8 and 41.1 respectively). Two breast fibroadenomas (dot-blot positions B4 and B6) contain very low levels of hybridizing RNA. Of all benign breast tissues analyzed to date, only one sample that was pathologically classified as nonmalignant (BN9-A7 on the dot blot) contains significant levels of 3b cDNA hybridizing RNA. Interestingly, this sample was obtained from the second breast of a breast cancer patient who had undergone mastectomy several years earlier. In contrast, mostly low, albeit detectable, levels of hybridization to the 3b cDNA probe are present in RNA extracted from stomach, colon and lung adenocarcinomas, as well as hypernephroma. Extremely low levels are seen in RNA isolated from a bladder carcinoma and undetectable levels of hybridization occurred with RNA from an adrenal pheochromocytoma, as well as in RNA extracted from chronic lymphocytic leukemic cells or from a brain neuroblastoma sample (data not shown).

Analyses of human tumor RNA species by Northern blotting

In order to determine by an independent method the validity of the dot-blot analysis, the human-tissue RNAs were analyzed by probing Northern blots with the 3b cDNA probe (Fig. 5). Differences in both the sizes of the hybridizing mRNA species as well as in the relative levels are immediately evident. The relative levels obtained in the Northern blot analysis correlate well with those seen in the initial dot-blot screening. The most intensive hybridization is observed with RNA extracted from the breast tumor BT15 which yields a prominent RNA band located at the 6.5-kb position along with a significantly weaker band at approximately 3.6 kb (Fig. 5B, lane 7). A densitometric analysis indicates that hybridization in this breast-tumor RNA sample is 30–40-fold higher than that observed in other RNA samples analyzed on the same Northern blot. Much lower levels of hybridization of a 3.6-kb species are seen in RNA isolated from the two pleural effusions (lanes 5 and 9) which were revealed by longer exposure of the autoradiograph (data not shown).

RNA extracted from the breast fibroadenoma (lane 6, Fig. 5B) demonstrates only very low levels of hybridization with a 3.0-kb RNA species, whereas barely detectable levels

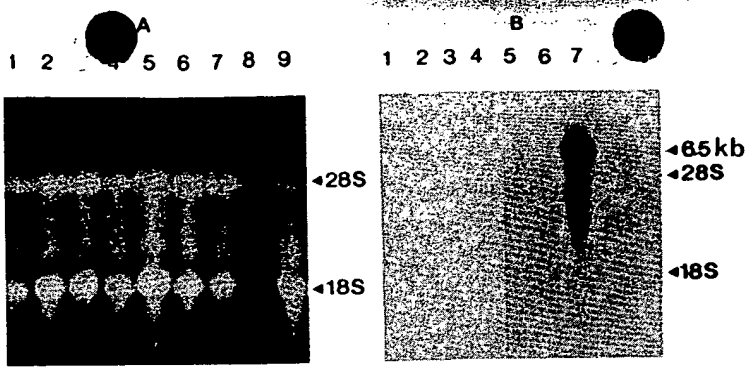


Fig. 5. *Northern blot analysis of RNA species hybridizing with 3b cDNA.* RNA samples isolated from a human hypernephroma KT1 (lane 1), adjacent kidney non-malignant tissue KN1 (lane 2), thyroid nodular goiter (lane 3), thyroid follicular adenoma and lymphocytic thyroiditis (lane 4), cells from the pleural effusions of two patients with advanced breast carcinoma PE1 and PE2 (lanes 5, 9), breast fibroadenoma BF16 (lane 6), breast adenocarcinoma BT15 (lane 7) and brain neuroblastoma (lane 8) were analyzed by agarose gel electrophoresis and ethidium bromide staining (A) followed by Northern blotting and probing with the 3b cDNA insert. (B) The autoradiograph of 1-day exposure is presented. Total RNA was extracted from tissue samples and analyzed by Northern blotting and probing with the 3b cDNA insert as described in Materials and Methods. All washings were performed under stringent conditions (0.2 × NaCl/Cit, twice for 30 min at 60 °C)

of hybridization at the 3.6-kb position are seen in one of the thyroid samples (lane 4, longer exposure, data not shown). Hybridization is not detected with RNA from a neuroblastoma, one thyroid sample and non-malignant tissue of the hypernephroma (lanes 8, 3 and 2, respectively).

The gene hybridizing with the repeat unit is polymorphic and is a VNTR gene: verification with unique non-repeat genomic sequences

The presence of the 60-nucleotide tandem repeat unit in the cDNAs analyzed indicates that the gene coding for this protein probably also contains a variable number of tandem repeats and thus belongs to the class known as VNTR genes. In order to demonstrate the polymorphism occurring in such a gene, a Southern blot comprising *Eco*RI and *Eco*RI/*Pst*I double-digested DNA was prepared from a number of human tissue samples isolated from different individuals. Hybridization with either the 3b cDNA insert or with the larger 850-bp cDNA insert (previously described above) shows marked gene polymorphism with at least 11 different alleles evident in the 9 samples studied (Fig. 6). Although the allelic patterns are similar on the *Eco*RI or double-digested *Eco*RI/*Pst*I DNA samples, the sizes of the different alleles following the double digestion are significantly smaller, thus increasing their electrophoretic resolution (Fig. 6).

From the Southern blot and cDNA nucleotide sequencing data presented, it is concluded that (a) the different alleles result from differences in the number of repeat units, (b) the *Eco*RI and *Pst*I sites are situated outside the tandem repeat unit and (c) the *Pst*I sites are closer to the borders of the tandem repeat units than are the *Eco*RI sites.

Polymorphism of this gene has also recently been described by two other groups [13–16]. However, the only probe used in the reported studies has been the 60-bp repeat unit [13–16]. Conclusive evidence that the gene is in fact an expressed VNTR gene requires probing of both Northern and Southern blots also with unique non-repeat sequences that are linked to the repeat array.

We further verified the VNTR nature of the gene by re-probing the same Southern blot with a non-repeat DNA fragment excised from the cloned 7.5-kb *Eco*RI–*Eco*RI gene fragment, isolated from a genomic library by probing with the cDNA 60-bp repeat unit [34]. This fragment (a *Sma*I–*Pst*I fragment, see Fig. 6), is approximately 1 kb and is situated 5′ to the array of tandem repeat units. It should thus hybridise with a single identically sized DNA band in all samples that have been *Eco*RI/*Pst*I double-digested. On the other hand, hybridization of this same fragment with an *Eco*RI-digested DNA should yield an identical hybridization pattern to that seen following hybridization with the repeat unit. These predictions were confirmed by the results obtained. Hybridization of the *Eco*RI/*Pst*I-digested samples with the 1-kb non-repeat fragment reveals a 3.5-kb band in all samples investigated (Fig. 6B, *Eco*RI + *Pst*I). This band is absent when the blot is probed with the repeat unit (compare Fig. 6A and B *Eco*RI + *Pst*I digest). The lightly labelled additional bands designated by asterisks in Fig. 6B are the remnants of the first hybridization with the repeat unit (compare with Fig. 6A, *Eco*RI + *Pst*I). Furthermore, hybridization of the *Eco*RI digest with the non-repeat fragment or with the repeat unit are identical, as predicted (compare Fig. 6A and B *Eco*RI).

As expected, the larger allele that contains more repeat units than the smaller allele shows a stronger signal following hybridization with the repeat unit probe (see Fig. 6, lane 2 for example).

The above data present evidence that the gene coding for the tumor-associated antigen is indeed a VNTR gene.

The different alleles are codominantly transcribed into corresponding mRNA species

We had previously seen significant heterogeneity in mRNA species that hybridize with 3b cDNA expressed in tumor samples isolated from different individuals. Two, and less often, only one hybridizing RNA band(s) are observed in any individual sample. As the gene itself is highly polymorphic, we investigated whether a correlation exists between the different allelic forms and the number and sizes of hybridizing RNAs expressed (Fig. 7).

Although, as noted above, 3b-hybridizing mRNAs are highly over-expressed in most malignant breast tissues, RNA isolated from other epithelial tumors also demonstrate hybridizing mRNA species albeit at lower levels. In order to establish the scope of this possible allele/mRNA correlation, investigations were performed both on non-breast and breast tumor

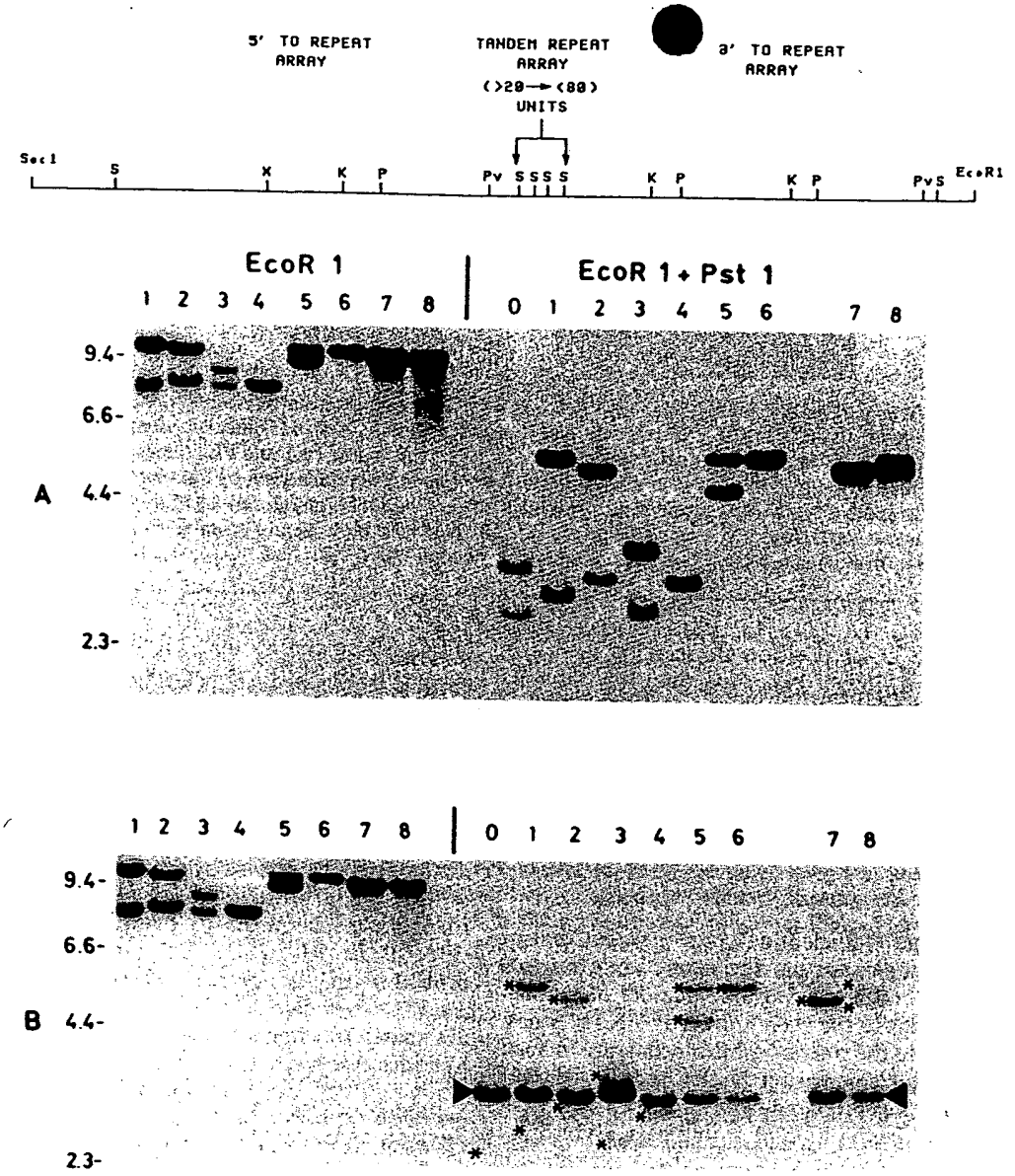


Fig. 6. *Hybridization of Southern blots with the repeating unit demonstrates a highly polymorphic gene.* High-molecular-mass genomic DNA was extracted from the following human organs that had malignant tumors: stomach (lane 0), ovary (lanes 1, 2), lung (lane 3), breast (lanes 4 and 6), colon (lane 5) and thyroid (lanes 7 and 8). The DNA was restricted with *Eco*RI alone or doubly digested with *Eco*RI and *Pst*I. (A) 10 µg was electrophoresed on agarose gels, Southern blotted and probed with radioactively labelled 850-bp cDNA. (B) Following this hybridization, the blot was rehybridized with a 1-kb non-repeat fragment of the gene (restriction map of gene, top panel, *Sma*I – *Pst*I fragment 5′ to the repeat array). The restriction enzymes *Kpn*I, *Pst*, *Pvu*II, *Sma*I and *Xmn*I are represented by K, P, Pv, S and X respectively. The blots were stringently washed and autoradiographed at −70°C. The bands labelled in B with the asterisk are the remaining signals of those seen in the previous hybridization with the repeat unit, whereas the specifically labelled band is shown by the full arrow (B, *Eco*RI + *Pst*I). The numbers to the left of the figure indicate size (kb) of markers

samples. Two breast tumor samples that express the lowest levels of 3b-hybridizing mRNA out of all malignant breast tissues analyzed were selected for comparison. In this regard it should be emphasized that the investigation was performed on nucleic acids isolated from primary human tissues rather than from cell culture lines. The conclusions of these experiments are therefore relevant to the *in-vivo* situation. In 10 out of 10 primary human tumor samples investigated, full concordance is demonstrated between the number and sizes of alleles with the corresponding hybridizing mRNA species (Fig. 7A and B).

The different allelic forms probably vary due to a difference in the number of tandem repeats. We thus investigated whether the corresponding mRNA species expressed in the same individual demonstrate an identical size difference. As the homozygotic breast tumor samples correspondingly express one mRNA species, they were not included in this analysis. The results shown in Fig. 7 indicate that, within the accuracy possible for DNA fragment and mRNA species size determination, the allelic size difference for the heterozygotic samples is equal to the difference in size of the two mRNA species.

It is interesting to note that the mRNA species correlating with the larger allele gives a less intense hybridization signal than the smaller mRNA species (see Fig. 7, lanes 2, 3 and 8). We do not know whether this is a consequence of reduced

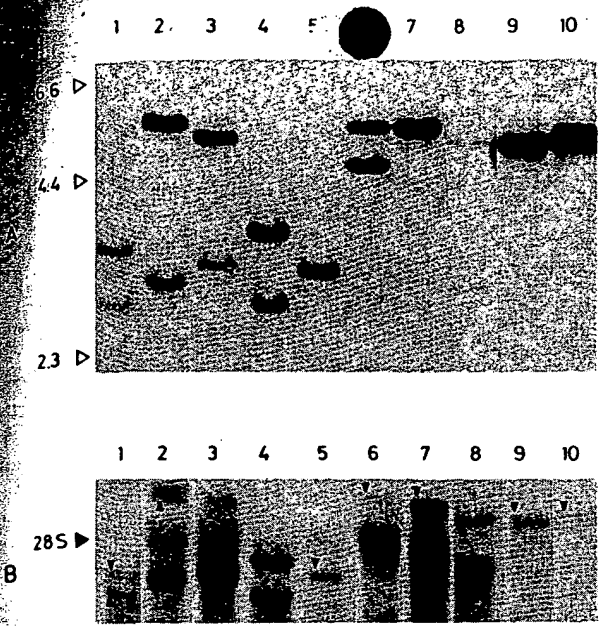


Fig. 7. *Correlation of Southern and Northern blots containing DNA and RNA isolated from the same human tissue sample following hybridization with the repeating unit.* DNA (A) or total RNA (B) was isolated from the following human tissues that had malignant tumors: stomach (lane 1), ovary (lanes 2 and 3), lung (lane 4), breast (lanes 5 and 7) and thyroid (lanes 8–10). The DNA after double digestion with *Eco*RI and *Pst*I or total RNA samples were electrophoresed on agarose gels and Southern (A) or Northern (B) blotted followed by hybridization with radioactively labelled 3b cDNA probe. The blots were stringently washed and autoradiographed at −70 °C. All samples in A were run simultaneously on the same gel but lane 8 was exposed for a longer time as less DNA was available for analysis. On the Northern blot, samples 2–7 were run simultaneously on the same gel but lane 6 was exposed for a longer time as there was significantly less mRNA expression in this sample. Samples 1 and samples 8–10 were run on two separate gels. The positions of the hybridizing mRNA species are indicated in B by the upward or downward facing full arrows. Note that in lanes 2–7 (and especially in lane 6) some nonspecific hybridization with 28S rRNA has occurred

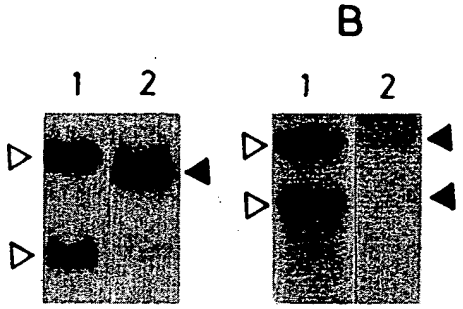


Fig. 8. *Probing Southern and Northern blots with the repeating-unit cDNA probe and a non-repeat genomic fragment.* High-molecular-mass DNA (A) or total RNA (B) was isolated from human lung tissue. The DNA, following *Eco*RI plus *Pst*I double digestion, and total RNA were electrophoresed on agarose gels and Southern (A) or Northern blotted (B). The blots were hybridized with the 3b repeating-unit cDNA probe (lane 1) or, after stripping, with the 1-kb non-repeat fragment of the gene (lane 2, see text). The blots were stringently washed and autoradiographed at −70 °C. The full arrows indicate bands hybridizing with the 1-kb non-repeat genomic fragment whilst the open arrows show bands hybridizing to the 3b repeat-unit cDNA probe. The efficiency of stripping the Northern blot following the first hybridization with 3b cDNA (B, lane 1) was evaluated by blot autoradiography prior to the second hybridization: no signal at all was seen. The bands appearing in B, lane 2, are thus *bona fide* signals

transcription of the larger allele, reduced stability of the larger mRNA species or other mechanisms.

In order to characterize further the correlation of allelic forms with the different mRNA species, both a Southern and Northern blot were rehybridized with the 1-kb non-repeat genomic fragment described above (Fig. 8A and B). As expected, probing the Northern blot with either the 3b cDNA tandem repeat units or with the 1-kb non-repeat fragment (Fig. 8B, lanes 1 and 2 respectively) reveals identical hybridizing mRNA species. On the other hand, reprobing the Southern blot with the 1-kb non-repeat fragment demonstrates only one band in contrast to the two allelic forms seen following probing with the repeat units (Fig. 8A, lanes 2 and 1 respectively).

Expression of H23 Ag in cells stably transfected with the H23 Ag gene

By probing Northern and Southern blots with both unique genomic sequences and the 60-bp repeat unit, we demonstrated the expression of a VNTR gene that codes for the H23 Ag. These critical experiments hinge on the physical linkage, in the genomic fragment isolated, of unique non-repeat DNA sequences with the tandem repeat array.

In order to confirm this linkage, mouse or rat cells were transfected with the isolated genomic fragment and then analyzed for H23 Ag synthesis. We had previously determined by cDNA and genomic sequencing (unpublished results) that an *Xmn*I site is located 35 nucleotides upstream to the putative ATG initiation codon of the H23 Ag gene. The *Xmn*I – *Eco*RI gene fragment (see Fig. 6) was therefore isolated and inserted into a eukaryotic expression vector downstream to the promoter of a housekeeping gene, 3-hydroxy-3-methylglutaryl-coenzyme-A reductase. In order to obtain stable transfectants, the H23 Ag gene construct, pCL642/*Xmn-Eco*, was cotransfected into mouse mammary tumor cells MM5, with a plasmid coding for resistance to the antibiotic neomycin. Similar transfections were conducted with c-Ha-ras-transformed rat fibroblasts. Neither of these cell lines expressed any human epithelial tumor antigen detectable with H23 mAb. As a control, MM5 cells were separately stably transfected with a pCL642 construct containing an irrelevant gene. (Details on the pCL642 eukaryotic expression vector are to be presented in a separate publication.)

Both the MM5 and rat fibroblast stable transfectants were grown on coverslips and immunohistochemically stained with H23 mAb (Fig. 9). Whereas no staining is observed in control MM5 and rat fibroblasts transfected with the non-relevant gene (Fig. 9A′ and B′), stable transfectants harboring the pCL642/*Xmn-Eco* construct demonstrate intense staining, readily detected with the H23 mAb (Fig. 9A and B). Staining is mainly cytoplasmic and is undetectable within the cell nucleus.

Western blot analyses of cell proteins from the pCL642/*Xmn-Eco* transfection demonstrate high-molecular-mass proteins (five bands ranging from 70 kDa to > 200 kDa) that are immunoreactive with H23 mAb (data not shown). These protein species are likely to represent H23 Ag glycosylated to varying degrees, thereby producing heterogenously sized immunoreactive products. Cell extracts from cells transfected with the non-relevant gene show no immunoreactive bands on the Western blot analysis.

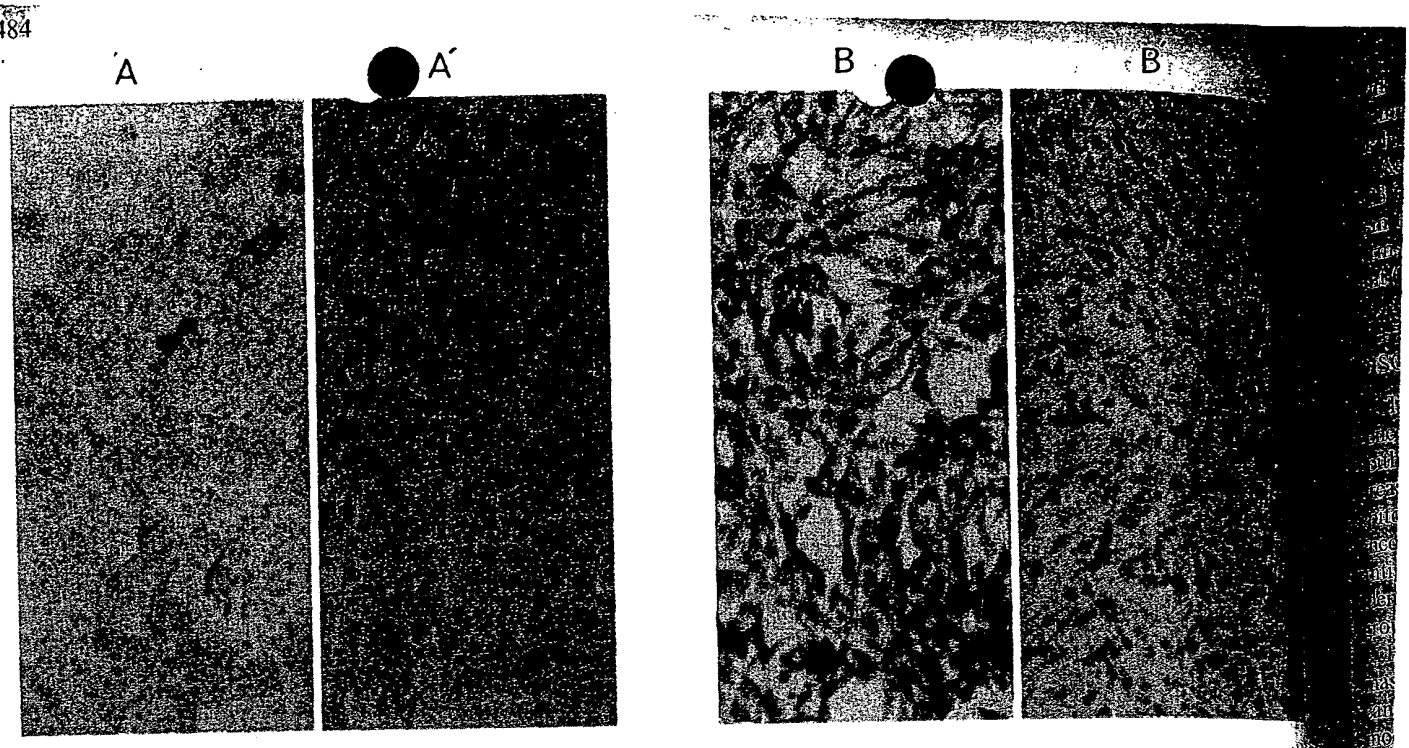


Fig. 9. *Expression of H23 Ag in cells transfected with the H23 Ag gene.* Mouse mammary tumor cells MM5 (A) or cHa-ras transformed fibroblasts (B) were transfected with the H23 Ag gene inserted into an expression vector, as described in Methods, grown on coverslips and immunohistochemically stained with H23 mAbs. Controls were MM5 cells (A') or cHa-ras transformed fibroblasts (B') transfected with a non-relevant gene and stained with H23 mAb. Intense cytoplasmic and membrane-bound staining is observed in the H23 Ag gene-transfected cells (A and B)

Over-expression of the H23 Ag in primary human breast tumor tissue: Western blot analysis

Having established (a) that in primary human tissues the gene polymorphism directly correlates with the mRNA species expressed and (b) that the mRNA coding for the antigen is over-expressed in breast tumor tissue, we next investigated the expression of antigen at the protein level.

A preliminary investigation was conducted on the human breast cell line T47D, which expresses large amounts of tumor antigen that are readily detectable by H23 mAbs. These cells were analysed at the gene, mRNA and protein levels.

A Southern blot of an *Eco*RI/*Pst*I digest shows two allelic forms at 5.5 and 3.1 kb (Fig. 10A, lane s). The Northern blot analysis correspondingly demonstrates two mRNA species (6.5 and 4.1 kb) that hybridize with the repeat-unit cDNA probe (data not shown). The protein products of these mRNA species were analyzed by immunoblotting which shows two products migrating on SDS-denaturing gels with molecular masses in the region of 250–450 kDa (Fig. 10A, lane w). No bands are observed when the immunoblot was probed with a non-specific monoclonal antibody under identical conditions. The alleles of the T47D gene are thus transcribed into mRNA species that are subsequently translated into distinct high-molecular-mass protein products that correlate with the respective mRNA and allelic sizes.

In order to relate the above findings to an *in-vivo* system, these studies were extended, as with the RNA and DNA analyses, to primary human tissues. Extracts of human tissue samples were run on SDS-denaturing gels and the separated protein species immunoblotted and probed with H23 mAbs. Analyses were performed on malignant breast tumor tissue samples together with an extract from non-malignant breast tissue adjacent to the biopsied tumor sample.

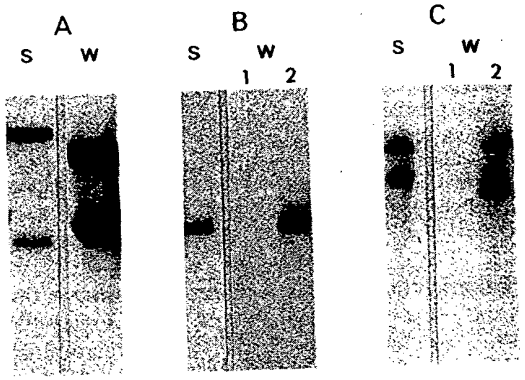


Fig. 10. *Correlation of alleles hybridizing to the repeat unit with protein products detected by H23 monoclonal antibodies and H23 Ag overexpression in breast cancer tissue.* DNA, double-digested with *Eco*RI and *Pst*I, or RNA isolated from T47D breast carcinoma cells were analyzed by agarose gel electrophoresis and Southern (A, s) or Northern blotted (not shown). The blots were hybridized to the 3b cDNA repeating unit probe, stringently washed and autoradiographed. For immunoblotting (A, B and C, w), samples were boiled for 3 min in SDS/mercaptoethanol sample buffer, and electrophoresed on a 4–15% SDS gradient gel, followed by electro-transfer to a nitrocellulose membrane as in Methods, and reacted with H23 mAb. Sample A, lane w is the medium of T47D cells precipitated with 50% ammonium sulfate; samples B and C are the protein extracts from breast cancer tissue (lane 2) and the adjacent non-malignant breast tissue (lane 1). DNA extracted from the same samples (B and C, lanes s) was restricted with *Eco*RI alone, Southern blotted and probed with 3b cDNA

Probing the immunoblots with H23 mAbs demonstrates marked over-expression of the tumor antigen in the malignant breast tissue samples (Fig. 10B and C, lane 2). The non-malig-

breast tissues, that were adjac[illegible] their malignant counterparts, show significantly lower immune reactivity with the H23 mAbs (Fig. 10B and C, lane 1).

As previously shown for T47D, the polymorphism of the H23 immunoreactive protein species seen in the primary human breast tissue samples correlates with the different allelic forms observed in Southern blots probed with the 60-bp repeat unit (Fig. 10B and C, lane s).

DISCUSSION

The results presented here show that a highly polymorphic gene contains a 60-bp tandem repeat array and codes for an epithelial tumor antigen that is over-expressed in human breast cancer. The H23 monoclonal antibody recognizes an epitope contained within the 20-amino-acid repeat motif encoded by the 60-bp cDNA and detects intracytoplasmic antigen in 91% of malignant breast tumors [2]. An almost identical 60-bp cDNA insert has been isolated by two other groups [13–16] using monoclonal antibodies (DF3, HMFG-1, HMFG-2 and SM3) that also recognize high-molecular-mass mucin-like glycoproteins aberrantly expressed in breast cancer tissue. It seems likely that different post-translational modifications occur within the 20-amino-acid repeat motif, encoded by the 60-bp cDNA, thus explaining, in part, the varying specificities of the different mAbs for normal and malignant breast tissue.

The gene coding for tumor antigen

As previously reported [13–16], the gene coding for the tumor antigen contains a variable number of tandem repeats and is highly polymorphic. We have extended this finding and probed restricted genomic DNA samples with unique non-repeat sequences isolated from a genomic fragment that contains the tandem repeat array. This analysis demonstrates that, external to the repeat array, the gene does not exhibit any heterogeneity, thereby indicating that the genetic polymorphism is solely due to varying numbers of the 60-bp tandem repeats. It is also demonstrated here that, besides expression of the 60-bp repeat units, unique non-repeat genomic sequences are expressed into mRNA and translated into protein.

The physical linkage of unique non-repeat sequences with the expressed 60-bp tandem repeat array was further confirmed by transfection experiments. The isolated gene fragment, from which the unique repeat sequences are derived, was transfected into mouse and rat cells that do not normally express any tumor antigen detectable with H23 mAb. The transfectants thus obtained synthesize human tumor antigen that is readily detected by H23 mAb. Furthermore, these transfection studies provide strong evidence that the isolated cDNA and gene fragment are indeed *bona fide* sequences that code for the human epithelial tumor antigen.

Correlation of alleles with expressed mRNA species and protein products: studies with primary human tissues

In a recent study involving only material derived from cells grown in culture [13], the gene polymorphism was found to correlate with both the mRNA species and protein forms detected. The different protein species observed in human urine by immunoblot analysis [16] also correlate with the various alleles. To our knowledge, there have been no reports demonstrating a concordance between the various alleles, mRNA species and p[illegible]rms in primary human tissues. We show here that in primary human tumors full concordance exists between the alleles and the transcribed mRNA species. This is demonstrated for nucleic acids extracted from breast, ovary, lung, stomach and thyroid tissues. Furthermore, it is shown that the allelic and mRNA size differences are equivalent in every sample of primary human tissue analyzed. These studies indicate that the heterogeneity in mRNA species is also solely due to the number of tandem repeats that they contain.

The correlative study of alleles and mRNA species in the same samples allows us to determine that approximately 1.9 kb in any individual mRNA species is represented by non-repeat sequences. The coding capacity of the tandem array is thus probably greater than 50% of the total protein, even in the smallest mRNA observed, and could code for more than 65% in the larger mRNA species.

Analyses of RNA samples from primary benign and malignant tumors demonstrate undetectable levels of hybridization in tissues of nonepithelial origin, whereas several non-mammary epithelial adenocarcinoma tumors display low levels of hybridization. However, RNA extracted from three ovarian carcinomas shows significant levels of hybridization with the 3b cDNA (an example of the intensity of hybridization is shown in Fig. 4A, dot-blot position E2). A question of obvious interest is whether this expression is due to the endocrine nature of these tissues.

The highest levels of hybridizing mRNA species are detected in malignant breast tumors. Non-malignant 'normal' tissue adjacent to the breast tumor samples, as well as non-malignant breast fibroadenomas, display much lower hybridization levels. The increased expression of the mRNA species hybridizing with the 3b cDNA probe thus strongly correlates with the malignant phenotype of the breast tissue. Although the mechanisms involved in the increased expression are not known, they may be related to the de-differentiated state of malignant tissue.

Since H23 mAbs detect an intracellular antigen primarily in breast tumor sections [2], the detection of hybridizing RNA from non-breast tumors with the 3b cDNA probe, albeit at low levels, is surprising. We have recently isolated unique-sequence cDNA that account for almost full-length cDNA of the tumor antigen [16a]. As several different alternatively spliced cDNAs were characterized, it is possible that the loss of epitope recognition by H23 mAb may be due to alternative splicing of the mRNA species in non-breast tissues. Other possibilities to be considered are different translational frames or simply a question of sensitivity of the immunohistochemical staining technique.

The expression of the gene coding for the tumor antigen was also investigated at the protein level. This study shows that in primary human breast tumor tissue the polymorphism detected in the gene and mRNA species correlates with the protein products detected by immunoblotting. Moreover, it is quite obvious that the malignant breast tissue contains significantly higher levels of tumor antigen than adjacent normal breast tissue.

We and others [12] have described a 68-kDa protein species that can be precipitated with mAbs directed against this epithelial tumor antigen. The 68-kDa protein is not detected, however, using the conditions of the immunoblot technique described here. It may represent a partially glycosylated protein or alternatively a breakdown product (induced proteolytically or otherwise), that contains a discrete number of

repeat motifs. These possibilities are presently being investigated.

The 20-amino-acid repeat motif: comparison with flanking amino acids

The epithelial tumor antigen is composed of 20-amino-acid repeats that make up more than 50% of the total protein. This unusual structure of highly conserved repeat motifs has recently been documented for porcine submaxillary gland mucin [28], human intestinal mucin [29], a cell-surface antigen expressed by murine hemopoietic progenitor cells [30], human apolipoprotein (a) [31], apo-polysialoglycoprotein of rainbow trout eggs [32] and a repetitive protein from *Xenopus laevis* skin [33]. It is interesting to note that repeat elements from several of these proteins are also rich in serine and threonine residues. The function of the 20-amino-acid repeat motif, as of the complete epithelial tumor antigen itself, is unknown. It is striking, however, that specific amino acids and subregions of the repeat element are conserved in the flanking regions on both sides of the repeat array. Furthermore, in some cases, identical amino acids replace a repeat motif amino acid in the same position on both sides of the repeat array. Although we do not understand the significance of this gradual decline in similarity between flanking amino acids and the repeat motif, it may indicate that a specific function is related to a certain amino acid sequence of the repeat motif.

Mouse cell transfectants synthesize the human epithelial tumor antigen: possible insights into tumor antigen function

Mouse cells transfected with the isolated gene coding for the human epithelial tumor antigen synthesize protein readily detected with H23 mAb. The location of synthesized antigen is primarily cytoplasmic, although we cannot rule out the possibility that it may be bound to the endoplasmic reticulum and/or plasma membrane. Recent analyses of full-length cDNAs [16a] show that differential splicing events occur 5′ and 3′ to the tandem repeat array giving mRNAs that will produce several forms of the antigen that localize to different cellular compartments.

Using the transfected cells as a model system, we are now in a position to ask questions regarding the function of this epithelial tumor antigen: does it change the growth characteristics, morphology or/and tumorgenicity of the transfected cells? Preliminary results indicate that expression of the tumor antigen indeed changes the cell growth potential; these and other possible functions are presently being investigated.

This work was supported by the Simko Chair for Breast Cancer Research, Frederico Fund for Tel Aviv University, grants from *Center National Récherche Scientifique*, Mr Toby Green (London), Maurice and Frida Eskanasy Cancer Research Fund and the Israel Cancer Association (DHW). DHW, MH and IT were the recipients of European Molecular Biology Organization short-term and long-term (MH) Fellowships; DHW is the recipient of a Koret Foundation Fellowship, California. We thank Professor Pierre Chambon for continued support and fruitful discussions.

REFERENCES

1. Keydar, I., Chen, L., Karby, S., Weiss, F. R., Delarea, J., Radu, M., Chaitchik, S. & Brenner, H. J. (1979) *Eur. J. Cancer 15*, 659–670.
2. Keydar, R., Chou, C. S., Hareuveni, M., Tsarfaty, I., Sahar, E., Seltzer, G., Chaitchik, S. & Hizi, A. (1989) *Proc. Natl Acad. Sci. USA 86*, 1362–1366.
3. Abe, M. & Kufe, D. (1986) *J. Cell. Physiol. 126*, 126–136.
4. Bramwell, M. E., Bhavanandan, V. P., Wiseman, G. & Ha… H. (1983) *Br. J. Cancer 48*, 177–183.
5. Burchell, J. M., Durbin, H. & Taylor-Papadimitriou, J. (1983) *Immunol. 131*, 508–513.
6. Ceriani, R. L., Peterson, J. A. & Blank, E. W. (1984) *Cancer R… 44*, 3033–3039.
7. Hilkens, J., Buijs, F., Hilgers, J., Hagemann, Ph., Calafat, … Sonnenberg, A. & Van der Valk, M. (1984) *Int. J. Cancer* … 197–206.
8. Johnson, V. G., Schlom, J., Paterson, A. J., Bennett, J., Magna… J. L. & Colcher, D. (1986) *Cancer Res. 46*, 850–857.
9. Lan, M. S., Finn, O. J., Fernsten, P. D. & Metzgar, R. S. (198… *Cancer Res. 45*, 305–310.
10. Magnani, J. L., Steplewski, Z., Koprowski, H. & Ginsburg, … (1983) *Cancer Res. 43*, 5489–5492.
11. Price, M. R., Edwards, S., Robins, R. A., Hilgers, J., Hilken… J. & Baldwin, R. (1986) *Eur. J. Can. Clin. Oncol. 22*, 115–11…
12. Burchell, J., Gendler, S., Taylor-Papadimitriou, J., Girling, A… Lewis, A., Mills, R. & Lamport, D. (1987) *Cancer Res. 4…* 5476–5482.
13. Siddiqui, J., Abe, M., Hayes, D., Shani, E., Yunis, E. & Kufe, D… (1988) *Proc. Natl Acad. Sci. USA 85*, 2320–2323.
14. Gendler, S., Taylor-Papadimitriou, J., Duhig, T., Rothbard, J. … Burchell, J. (1988) *J. Biol. Chem. 263*, 12820–12823.
15. Gendler, S. J., Burchell, J. M., Duhig, T., Lamport, D., White… R., Parker, M. & Taylor-Papadimitriou, J. (1987) *Proc. Natl … Acad. Sci. USA 84*, 6060–6064.
16. Swallow, D. M., Gendler, S., Griffiths, B., Corney, G., Taylor-… Papadimitriou, J. & Bramwell, M. E. (1987) *Nature 328*, 82–84.
16a. Wreschner, D. H., Hareuveni, M., Tsarfaty, I., Smorodinsky, N., Horev, J., Zaretsky, J., Kotkes, P., Weiss, M., Lathe, R., Dion, A. & Keydar, I. (1990) *Eur. J. Biochem. 189*, 463–473.
17. Petkovitch, M., Brand, N. J., Krust, A. & Chambon, P. (1987) *Nature 330*, 445–450.
18. Walter, P., Green, S., Green, G., Krust, A., Bornert, J. M., Jeltsch, J. M., Staub, A., Jensen, E., Scrace, G., Waterfield, M. & Chambon, P. (1985) *Proc. Natl Acad. Sci. USA 82*, 7889–7893.
19. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol. 113*, 237–251.
20. Wreschner, D. H. & Rechavi, G. (1988) *Eur. J. Biochem. 172*, 333–344.
21. Aviv, H. & Leder, P. (1972) *Proc. Natl Acad. Sci. USA 69*, 1408–1412.
22. Colbere-Garapin, F., Horodniceanu, F., Kourilsky, P. & Garapin, A. C. (1981) *J. Mol. Biol. 150*, 1–14.
23. Matriceau, L. M., Glaichenhaus, N., Gesnel, M. C. & Breathnach, R. (1985) *EMBO J. 4*, 1435–1440.
24. Wigler, S. F., Pellicer, A., Silverstein, S. & Axel, R. (1978) *Cell 14*, 725–731.
25. Graham, F. & Van der Eb, A. (1973) *Virology 52*, 456–457.
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl Acad. Sci. USA 74*, 5463–5467.
27. Laemmli, U. K. (1970) *Nature 227*, 680–685.
28. Timpte, C. S., Eckhardt, A. E., Abernethy, J. L. & Hill, R. L. (1988) *J. Biol. Chem. 263*, 1081–1088.
29. Gun, J. R., Byrd, J. C., Hicks, J. W., Toribara, J. W., Lamport, D. T. A. & Kim, Y. S. (1989) *J. Biol. Chem. 264*, 6480–6487.
30. Dougherty, G. J., Kay, R. J. & Humphries, R. K. (1989) *J. Biol. Chem. 264*, 6509–6514.
31. McLean, J. W., Tomlinson, J. E., Kuang, W. J., Eaton, D. L., Chen, E. Y., Fless, G. M., Scanu, A. M. & Lawan, R. M. (1987) *Nature 330*, 132–137.
32. Sorimachi, M., Emori, Y., Kawasaki, H., Kitajima, D., Inoue, S., Suzuki, K. & Inoue, Y. (1988) *J. Biol. Chem. 263*, 17678–17684.
33. Hoffman, W. (1988) *J. Biol. Chem. 263*, 7686–7690.
34. Tsarfaty, I., Hareuveni, M., Horev, J., Zaretsky, J., Weiss, M., Jeltsch, J. M., Garnier, J. M., Lathe, R., Keydar, I. & Wreschner, D. H. (1990) *Gene*, in the press.

1: J Parasitol. 2003 Apr;89(2):381-4. Related Articles, Lin

Upregulation of cardiac cell plasma membrane calcium pump in a canine model of Chagas disease.

Barr SC, Pannabecker TL, Gilmour RF Jr, Chandler JS.

Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA. scb6@cornell.edu

We have previously demonstrated that cardiac myocytes isolated from the hearts of adult dogs develop rapid repetitive cytosolic Ca2+ transients, membrane depolarization, and cell contraction by mobilization of sarcoplasmic reticulum Ca2+ stores when exposed to a soluble factor from th trypomastigotes of Trypanosoma cruzi. These findings led us to investigate the regulatory mechanisms of cytosolic Ca2+ in cardiac tissues from dogs chronically infected with T. cruzi. Expression of the plasma membrane calcium pump (PMCA) RNA and protein was determined by Northern and Western blotting, respectively, followed by densitometric analyses. A 642-bp PMCA 1b complementary DNA probe derived from canine epicardial tissue hybridized to 2 major transcripts (7.3 and 5.3 kb) in canine epicardium. Expression of the dominant transcript (7.3 kb) was 77% greater in cardiac tissues obtained from dogs with chronic T. cruzi infection (140 days after inoculation) in comparison with constitutive expression levels in normal dog Monoclonal antibody 5F10, known to recognize all isoforms of the PMCA, was used to detect expression of the PMCA protein in epicardial tissue. Expression of a 142-kDa protein was increased by 58% in the cardiac tissues of infected dogs when compared with those from uninfected dogs. To establish a link between the upregulation of PMCA in dogs chronically infected with Chagas disease and the ventricular-based arrhythmias and myocardial failure that occur during this stage of disease both in dogs and humans, further study will be required.

PMID: 12760659 [PubMed - indexed for MEDLINE]

Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions

JAMES U. BOWIE,* JOHN F. REIDHAAR-OLSON, WENDELL A. LIM, ROBERT T. SAUER

An amino acid sequence encodes a message that determines the shape and function of a protein. This message is highly degenerate in that many different sequences can code for proteins with essentially the same structure and activity. Comparison of different sequences with similar messages can reveal key features of the code and improve understanding of how a protein folds and how it performs its function.

THE GENOME IS MANIFEST LARGELY IN THE SET OF PROteins that it encodes. It is the ability of these proteins to fold into unique three-dimensional structures that allows them to function and carry out the instructions of the genome. Thus, comprehending the rules that relate amino acid sequence to structure is fundamental to an understanding of biological processes. Because an amino acid sequence contains all of the information necessary to determine the structure of a protein (*1*), it should be possible to predict structure from sequence, and subsequently to infer detailed aspects of function from the structure. However, both problems are extremely complex, and it seems unlikely that either will be solved in an exact manner in the near future. It may be possible to obtain approximate solutions by using experimental data to simplify the problem. In this article, we describe how an analysis of allowed amino acid substitutions in proteins can be used to reduce the complexity of sequences and reveal important aspects of structure and function.

Methods for Studying Tolerance to Sequence Variation

There are two main approaches to studying the tolerance of an amino acid sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. This method has been extremely powerful for proteins such as the globins or cytochromes, for which sequences from many different species are known (*2–7*). The second approach uses genetic methods to introduce amino acid changes at specific positions in a cloned gene and uses selections or screens to identify functional sequences. This approach has been used to great advantage for proteins that can be expressed in bacteria or yeast, where the appropriate genetic manipulations are possible (*3, 8–11*). The end results of both methods are lists of active sequences that can be compared and analyzed to identify sequence features that are essential for folding or function. If a particular property of a side chain, such as charge or size, is important at a given position, only side chains that have the required property will be allowed. Conversely, if the chemical identity of the side chain is unimportant, then many different substitutions will be permitted.

Studies in which these methods were used have revealed that proteins are surprisingly tolerant of amino acid substitutions (*2–4, 11*). For example, in studying the effects of approximately 1500 single amino acid substitutions at 142 positions in *lac* repressor, Miller and co-workers found that about one-half of all substitutions were phenotypically silent (*11*). At some positions, many different, nonconservative substitutions were allowed. Such residue positions play little or no role in structure and function. At other positions, no substitutions or only conservative substitutions were allowed. These residues are the most important for *lac* repressor activity.

What roles do invariant and conserved side chains play in proteins? Residues that are directly involved in protein functions such as binding or catalysis will certainly be among the most conserved. For example, replacing the Asp in the catalytic triad of trypsin with Asn results in a 10^4-fold reduction in activity (*12*). A similar loss of activity occurs in λ repressor when a DNA binding residue is changed from Asn to Asp (*13*). To carry out their function, however, these catalytic residues and binding residues must be precisely oriented in three dimensions. Consequently, mutations in residues that are required for structure formation or stability can also have dramatic effects on activity (*10, 14–16*). Hence, many of the residues that are conserved in sets of related sequences play structural roles.

Substitutions at Surface and Buried Positions

In their initial comparisons of the globin sequences, Perutz and co-workers found that most buried residues require nonpolar side chains, whereas few features of surface side chains are generally conserved (*6*). Similar results have been seen for a number of protein families (*2, 4, 5, 7, 17, 18*). An example of the sequence tolerance at surface versus buried sites can be seen in Fig. 1, which shows the allowed substitutions in λ repressor at residue positions that are near the dimer interface but distant from the DNA binding surface of the

The authors are in the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

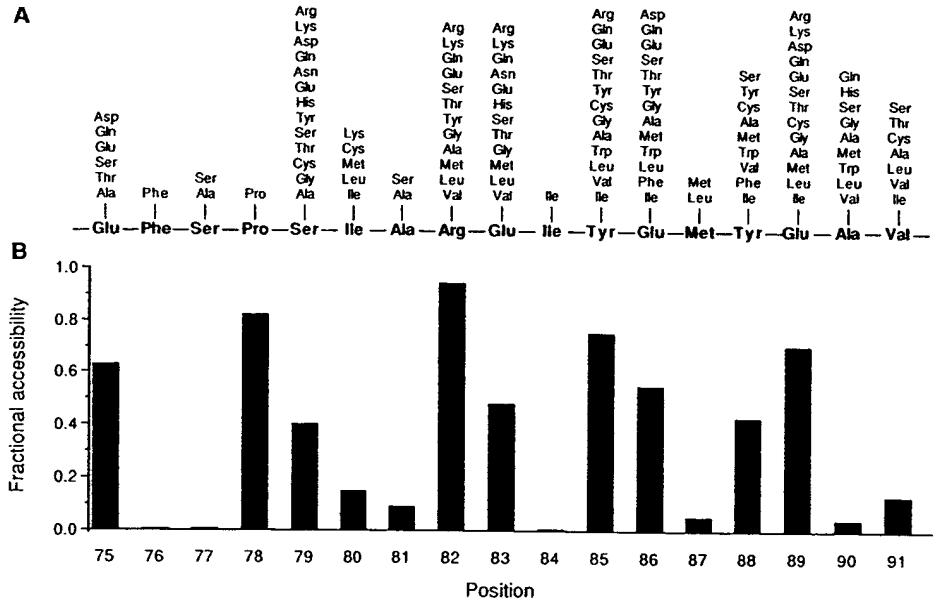


Fig. 1. (**A**) Amino acid substitutions allowed in a core region of λ repressor. The wild-type sequence is shown along the center line. The allowed substitutions shown above each position were identified by randomly mutating one to three codons at a time by using a cassette method and applying a functional selection (*9*). (**B**) The fractional solvent accessibility (*42*) of the wild-type side chain in the protein dimer (*43*) relative to the same atoms in an Ala-X-Ala model tripeptide.

selection after cassette mutagenesis. A histogram of side chain solvent accessibility in the crystal structure of the dimer is also shown in Fig. 1. At six positions, only the wild-type residue or relatively conservative substitutions are allowed. Five of these positions are buried in the protein. In contrast, most of the highly exposed positions tolerate a wide range of chemically different side chains, including hydrophilic and hydrophobic residues. Hence, it seems that most of the structural information in this region of the protein is carried by the residues that are solvent inaccessible.

Constraints on Core Sequences

Because core residue positions appear to be extremely important for protein folding or stability, we must understand the factors that dictate whether a given core sequence will be acceptable. In general, only hydrophobic or neutral residues are tolerated at buried sites in proteins, undoubtedly because of the large favorable contribution of the hydrophobic effect to protein stability (*19*). For example, Fig. 2 shows the results of genetic studies used to investigate the substitutions allowed at residue positions that form the hydrophobic core of the NH_2-terminal domain of λ repressor (*20*). The acceptable core sequences are composed almost exclusively of Ala, Cys, Thr, Val, Ile, Leu, Met, and Phe. The acceptability of many different residues at each core position presumably reflects the fact that the hydrophobic effect, unlike hydrogen bonding, does not depend on specific residue pairings. Although it is possible to imagine a hypothetical core structure that is stabilized exclusively by residues forming hydrogen bonds and salt bridges, such a core would probably be difficult to construct because hydrogen bonds require pairing of donors and acceptors in an exact geometry. Thus the repertoire of possible structures that use a polar core would probably be extremely limited (*21*). Polar and charged residues are occasionally found in the cores of proteins, but only at positions where their hydrogen bonding needs can be satisfied (*22*).

The cores of most proteins are quite closely packed (*23*), but some volume changes are acceptable. In λ repressor, the overall core volume of acceptable sequences can vary by about 10%. Changes at individual sites, however, can be considerably larger. For example, as shown in Fig. 2, both Phe and Ala are allowed at the same core position in the appropriate sequence contexts. Large volume changes

phylogenetic studies, where it has been noted that the size decreases and increases at interacting residues are not necessarily related in a simple complementary fashion (*5, 7, 17*). Rather, local volume changes are accommodated by conformational changes in nearby side chains and by a variety of backbone movements.

The Informational Importance of the Core

With occasional exceptions, the core must remain hydrophobic and maintain a reasonable packing density. However, since the core is composed of side chains that can assume only a limited number of conformations (*24*), efficient packing must be maintained without steric clashes. How important are hydrophobicity, volume, and steric complementarity in determining whether a given sequence can form an acceptable core? Each factor is essential in a physical sense, as a stable core is probably unable to tolerate unsatisfied hydrogen bonding groups, large holes, or steric overlaps (*25*). However, in an informational sense, these factors are not equivalent. For example, in experiments in which three core residues of λ repressor were mutated simultaneously, volume was a relatively unimportant informational constraint because three-quarters of all possible combinations of the 20 naturally occurring amino acids had volumes within the range tolerated in the core, and yet most of these sequences were unacceptable (*20*). In contrast, of the sequences that contained only

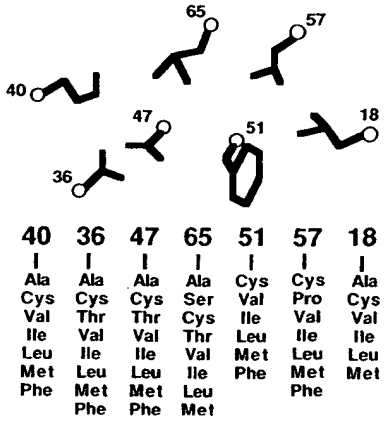


Fig. 2. Amino acid substitutions allowed in the core of λ repressor. The wild-type side chains are shown pictorially in the approximate orientation seen in the crystal structure (*43*). The lists of allowed substitutions at each position are shown below the wild-type side chains. These substitutions were identified by randomly mutating one to four residues at a time by using a cassette method and applying a functional selection (*20*). Not all substitutions are allowed in every sequence back-

the appropriate hydrophobic residues a significant fraction were acceptable. Hence, the hydrophobicity of a sequence contains more information about its potential acceptability in the core than does the total side chain volume. Steric compatibility was intermediate between volume and hydrophobicity in informational importance.

The Informational Importance of Surface Sites

We have noted that many surface sites can tolerate a wide variety of side chains, including hydrophilic and hydrophobic residues. This result might be taken to indicate that surface positions contain little structural information. However, Bashford *et al.*, in an extensive analysis of globin sequences (*4*), found a strong bias against large hydrophobic residues at many surface positions. At one level, this may reflect constraints imposed by protein solubility, because large patches of hydrophobic surface residues would presumably lead to aggregation. At a more fundamental level, protein folding requires a partitioning between surface and buried positions. Consequently, to achieve a unique native state without significant competition from other conformations, it may be important that some sites have a decided preference for exterior rather than interior positions. As a result, many surface sites can accept hydrophobic residues individually, but the surface as a whole can probably tolerate only a moderate number of hydrophobic side chains.

Identification of Residue Roles from Sets of Sequences

Often, a protein of interest is a member of a family of related sequences. What can we infer from the pattern of allowed substitutions at positions in sets of aligned sequences generated by genetic or phylogenetic methods? Residue positions that can accept a number of different side chains, including charged and highly polar residues, are almost certain to be on the protein surface. Residue positions that remain hydrophobic, whether variable or not, are likely to be buried within the structure. In Fig. 3, those residue positions in λ repressor that can accept hydrophilic side chains are shown in orange and those that cannot accept hydrophilic side chains are shown in green. The obligate hydrophobic positions define the core of the structure, whereas positions that can accept hydrophilic side chains define the surface.

Functionally important residues should be conserved in sets of active sequences, but it is not possible to decide whether a side chain is functionally or structurally important just because it is invariant or conserved. To make this distinction requires an independent assay of protein folding. The ability of a mutant protein to maintain a stably folded structure can often be measured by biophysical techniques, by susceptibility to intracellular proteolysis (*26*), or by binding to antibodies specific for the native structure (*27, 28*). In the latter cases, it is possible to screen proteins in mutated clones for the ability to fold even if these proteins are inactive. Sets of sequences that allow formation of a stable structure can then be compared to the sets that allow both folding and function, with the active site or binding residues being those that are variable in the set of stable proteins but invariant in the set of functional proteins. The DNA-binding residues of Arc repressor were identified by this method (*8*). The receptor-binding residues of human growth hormone were also identified by comparing the stabilities and activities of a set of mutant sequences (*28*). However, in this case, the mutants were generated as hybrid sequences between growth hormone and related hormones with different binding specificities.

Implications for Structure Prediction

At present, the only reliable method for predicting a low-resolution tertiary structure of a new protein is by identifying sequence similarity to a protein whose structure is already known (*29, 30*). However, it is often difficult to align sequences as the level of sequence similarity decreases, and it is sometimes impossible to detect statistically significant sequence similarity between distantly related proteins. Because the number of known sequences is far greater than the number of known structures, it would be advantageous to increase the reach of the available structural information by improving methods for detecting distant sequence relations and for subsequently aligning these sequences based on structural principles. In a normal homology search, the sequence database is scanned with a single test sequence, and every residue must be weighted equally. However, some residues are more important than others and should be weighted accordingly. Moreover, certain regions of the protein are more likely to contain gaps than others. Both kinds of information can be obtained from sequence sets, and several techniques have

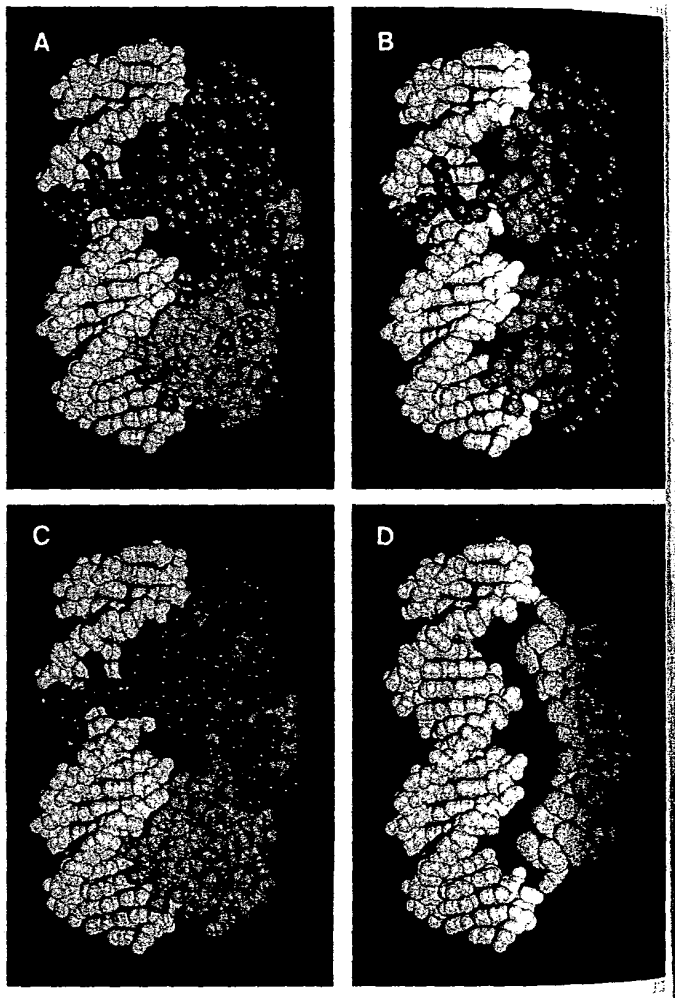


Fig. 3. Tolerance of positions in the NH_2-terminal domain of λ repressor to hydrophilic side chains. The complex (*43*) of the repressor dimer (blue) and operator DNA (white) is shown. In (**A**), positions that can tolerate hydrophilic side chains are shown in orange. The same side chains are shown in (**B**) without the remaining protein atoms. In (**C**), positions that require hydrophobic or neutral side chains are shown in green. These side chains are shown in (**D**) without the remaining protein atoms. About three-fourths of the 92 side chains in the NH_2-terminal domain are included in both (*B*) and (D). The remaining positions have not been tested. Data are from (*9, 14, 20, 27, 44*).

een used to combine such information more appropriately ighted sequence searches and alignments (*31*). These methods re used to align the sequences of retroviral proteases with aspartic oteases, which in turn allowed construction of a three-dimension- model for the protease of human immunodeficiency virus type 1 9). Comparison with the recently determined crystal structure of is protein revealed reasonable agreement in many areas of the edicted structure (*32*).

The structural information at most surface sites is highly degener- e. Except for functionally important residues, exterior positions em to be important chiefly in maintaining a reasonably polar rface. The information contained in buried residues is also egenerate, the main requirement being that these residues remain drophobic. Thus, at its most basic level, the key structural essage in an amino acid sequence may reside in its specific pattern hydrophobic and hydrophilic residues. This is meant in an formational sense. Clearly, the precise structure and stability of a otein depends on a large number of detailed interactions. It is ssible, however, that structural prediction at a more primitive vel can be accomplished by concentrating on the most basic formational aspects of an amino acid sequence. For example, mphipathic patterns can be extracted from aligned sets of sequences d used, in some cases, to identify secondary structures.

If a region of secondary structure is packed against the hydropho- core, a pattern of hydrophobic residues reflecting the periodicity the secondary structure is expected (*33, 34*). These patterns can be scured in individual sequences by hydrophobic residues on the otein surface. It is rare, however, for a surface position to remain drophobic over the course of evolution. Consequently, the am- pathic patterns expected for simple secondary structures can be ch clearer in a set of related sequences (*6*). This principle is strated in Fig. 4, which shows helical hydrophobic moment plots the Antennapedia homeodomain sequence (Fig. 4A) and for a mposite sequence derived from a set of homologous homeodo- in proteins (Fig. 4B) (*35*). The hydrophobic moment is a simple asure of the degree of amphipathic character of a sequence in a en secondary structure (*34*). The amphipathic character of the e α-helical regions in the Antennapedia protein (*36*) is clearly ealed only by the analysis of the combined set of homeodomain quences. The secondary structure of Arc repressor, a small DNA- ding protein, was recently predicted by a similar method (*8*) and nfirmed by nuclear magnetic resonance studies (*37*).

The specific pattern of hydrophobic and hydrophilic residues in amino acid sequence must limit the number of different structures ven sequence can adopt and may indeed define its overall fold. If is true, then the arrangement of hydrophobic and hydrophilic sidues should be a characteristic feature of a particular fold. Sweet d Eisenberg have shown that the correlation of the pattern of drophobicity between two protein sequences is a good criterion r their structural relatedness (*38*). In addition, several studies dicate that patterns of obligatory hydrophobic positions identified m aligned sequences are distinctive features of sequences that pt the same structure (*4, 29, 38, 39*). Thus, the order of drophobic and hydrophilic residues in a sequence may actually be fficient information to determine the basic folding pattern of a otein sequence.

Although the pattern of sequence hydrophobicity may be a aracteristic feature of a particular fold, it is not yet clear how such atterns could be used for prediction of structure de novo. It is portant to understand how patterns in sequence space can be lated to structures in conformation space. Lau and Dill have proached this problem by studying the properties of simple quences composed only of H (hydrophobic) and P (polar) groups n two-dimensional lattices (*40*). An example of such a representation is shown in Fig. Residues adjacent in the sequence must occupy adjacent squares on the lattice, and two residues cannot occupy the same space. Free energies of particular conformations are evaluated with a single term, an attraction of H groups. By considering chains of ten residues, an exhaustive conformational search for all 1024 possible sequences of H and P residues was possible. For longer sequences only a representative fraction of the allowed sequence or conformation space could be explored. The significant results were as follows: (i) not all sequences can fold into a "native" structure and only a few sequences form a unique native structure; (ii) the probability that a sequence will adopt a unique native structure increases with chain length; and (iii) the native states are compact, contain a hydrophobic core surrounded by polar residues, and contain significant secondary structure. Although the gap between these two-dimensional simulations and three-dimensional structures is large, the use of simple rules and sequence representations yields results similar to those expected for real proteins. Three-dimensional lattice methods are also beginning to be developed and evaluated (*41*).

Summary

There is more information in a set of related sequences than in a single sequence. A number of practical applications arise from an analysis of the tolerance of residue positions to change. First, such information permits the evaluation of a residue's importance to the function and stability of a protein. This ability to identify the essential elements of a protein sequence may improve our understanding of the determinants of protein folding and stability as well as protein function. Second, patterns of tolerance to amino acid substitutions of varying hydrophilicity can help to identify residues likely to be buried in a protein structure and those likely to occupy

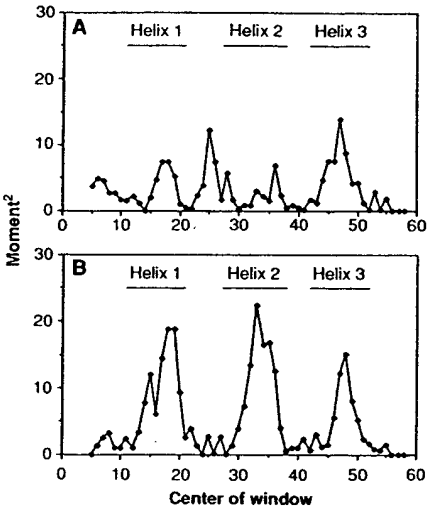


Fig. 4. Helical hydrophobic moments calculated by using (**A**) the Antennapedia homeodomain sequence or (**B**) a set of 39 aligned homeodomain sequences (*35*). The bars indicate the extent of the helical regions identified in nuclear magnetic resonance studies of the Antennapedia homeodomain (*36*). To determine hydrophobic moments, residues were assigned to one of three groups: H1 (high hydrophobicity = Trp, Ile, Phe, Leu, Met, Val, or Cys); H2 (medium hydrophobicity = Tyr, Pro, Ala, Thr, His, Gly, or Ser); and H3 (low hydrophobicity = Gln, Asn, Glu, Asp, Lys, or Arg). For the aligned homeodomain sequences, the residues at each position were sorted by their hydrophobicity by using the scale of Fauchere and Pliska (*45*). Arg and Lys were not counted unless no other residue was found at the position, because they contain long aliphatic side chains and can thereby substitute for nonpolar residues at some buried sites. To account for possible sequence errors and rare exceptions, the most hydrophilic residue allowed at each position was discarded unless it was observed twice. The second most hydrophilic residue was then chosen to represent the hydrophobicity of each position. An eight-residue window was used and the vectors projected radially every 100°. The vector magnitudes were assigned a value of 1, 0, or −1 for positions where the hydrophobicity group was H1, H2, or H3, respectively.

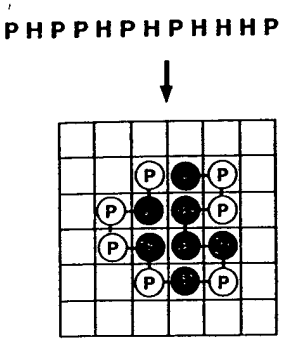


Fig. 5. A representation of one compact conformation for a particular sequence of H and P residues on a two-dimensional square lattice. [Adapted from (*40*), with permission of the American Chemical Society]

surface positions. The amphipathic patterns that emerge can be used to identify probable regions of secondary structure. Third, incorporating a knowledge of allowed substitutions can improve the ability to detect and align distantly related proteins because the essential residues can be given prominence in the alignment scoring.

As more sequences are determined, it becomes increasingly likely that a protein of interest is a member of a family of related sequences. If this is not the case, it is now possible to use genetic methods to generate lists of allowed amino acid substitutions. Consequently, at least in the short term, it may not be necessary to solve the folding problem for individual protein sequences. Instead, information from sequence sets could be used. Perhaps by simplifying sequence space through the identification of key residues, and by simplifying conformation space as in the lattice methods, it will be possible to develop algorithms to generate a limited number of trial structures. These trial structures could then, in turn, be evaluated by further experiments and more sophisticated energy calculations.

REFERENCES AND NOTES

1. C. J. Epstein, R. F. Goldberger, C. B. Anfinsen, *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439 (1963); C. B. Anfinsen, *Science* **181**, 223 (1973).
2. R. E. Dickerson, *Sci. Am.* **242**, 136 (March 1980).
3. M. D. Hampsey, G. Das, F. Sherman, *FEBS Lett.* **231**, 275 (1988).
4. D. Bashford, C. Chothia, A. M. Lesk, *J. Mol. Biol.* **196**, 199 (1987).
5. A. M. Lesk and C. Chothia, *ibid.* **136**, 225 (1980).
6. M. F. Perutz, J. C. Kendrew, H. C. Watson, *ibid.* **13**, 669 (1965).
7. C. Chothia and A. M. Lesk, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 399 (1965).
8. J. U. Bowie and R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2152 (1989).
9. J. F. Reidhaar-Olson and R. T. Sauer, *Science* **241**, 53 (1988); *Proteins Struct. Funct. Genet.*, in press.
10. D. Shortle, *J. Biol. Chem.* **264**, 5315 (1989).
11. J. H. Miller *et al.*, *J. Mol. Biol.* **131**, 191 (1979).
12. S. Sprang *et al.*, *Science* **237**, 905 (1987); C. S. Craik, S. Roczniak, C. Largman, W. J. Rutter, *ibid.*, p. 909.
13. H. C. M. Nelson and R. T. Sauer, *J. Mol. Biol.* **192**, 27 (1986).
14. M. H. Hecht, J. M. Sturtevant, R. T. Sauer, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5685 (1984).
15. T. Alber, D. Sun, J. A. Nye, D. C. Muchmore, B. W. Matthews, *Biochemistry* **26**, 3754 (1987).
16. D. Shortle and A. K. Meeker, *Proteins Struct. Funct. Genet.* **1**, 81 (1986).
17. A. M. Lesk and C. Chothia, *J. Mol. Biol.* **160**, 325 (1982).
18. W. R. Taylor, *ibid.* **188**, 233 (1986).
19. W. Kauzmann, *Adv. Protein Chem.* **14**, 1 (1959); R. L. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8069 (1986).
20. W. A. Lim and R. T. Sauer, *Nature* **339**, 31 (1989); in preparation.
21. Lesk and Chothia (*5*) have argued that a protein core composed solely of hydrogen-bonded residues would also be inviable on evolutionary grounds, as a mutational change in one core residue would require compensating changes in any interacting residue or residues to maintain a stable structure.
22. T. M. Gray and B. W. Matthews, *J. Mol. Biol.* **175**, 75 (1984); E. N. Baker and R. E. Hubbard, *Prog. Biophys. Mol. Biol.* **44**, 97 (1984).
23. F. M. Richards, *J. Mol. Biol.* **82**, 1 (1974).
24. J. W. Ponder and F. M. Richards, *ibid.* **193**, 775 (1987).
25. J. T. Kellis, Jr., K. Nyberg, A. R. Fersht, *Biochemistry* **28**, 4914 (1989); W. S. Sandberg and T. C. Terwilliger, *Science* **245**, 54 (1989).
26. A. A. Pakula and R. T. Sauer, *Proteins Struct. Funct. Genet.* **5**, 202 (1989).
27. B. C. Cunningham and J. A. Wells, *Science* **244**, 1081 (1989); R. M. Breyer and R. T. Sauer, *J. Biol. Chem.* **264**, 13348 (1989).
28. B. C. Cunningham, P. Jhurani, P. Ng, J. A. Wells, *Science* **243**, 1330 (1989).
29. L. H. Pearl and W. R. Taylor, *Nature* **329**, 351 (1987).
30. W. J. Brown *et al.*, *J. Mol. Biol.* **42**, 65 (1969); J. Greer, *ibid.* **153**, 1027 (1981); J. M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 99 (1988).
31. W. R. Taylor, *Protein Eng.* **2**, 77 (1988).
32. M. A. Navia *et al.*, *Nature* **337**, 615 (1989).
33. M. Schiffer and A. B. Edmundson, *Biophys. J.* **7**, 121 (1967); V. I. Lim, *J. Mol. Biol.* **88**, 857 (1974); *ibid.*, p. 873.
34. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Nature* **299**, 371 (1982); D. Eisenberg, D. Schwarz, M. Komaromy, R. Wall, *J. Mol. Biol.* **179**, 125 (1984); D. Eisenberg, R. M. Weiss, T. C. Terwilliger, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 140 (1984).
35. T. R. Burglin, *Cell* **53**, 339 (1988).
36. G. Otting *et al.*, *EMBO J.* **7**, 4305 (1988).
37. J. N. Breg, R. Boelens, A. V. E. George, R. Kaptein, *Biochemistry* **28**, 9826 (1989); M. G. Zagorski, J. U. Bowie, A. K. Vershon, R. T. Sauer, D. J. Patel, *ibid.*, p. 9813.
38. R. M. Sweet and D. Eisenberg, *J. Mol. Biol.* **171**, 479 (1983).
39. J. U. Bowie, N. D. Clarke, C. O. Pabo, R. T. Sauer, *Proteins Struct. Funct. Genet.*, in preparation.
40. K. F. Lau and K. A. Dill, *Macromolecules* **22**, 3986 (1989).
41. A. Sikorski and J. Skolnick, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2668 (1989); A. Kolinski, J. Skolnick, R. Yaris, *Biopolymers* **26**, 937 (1987); D. G. Covell and R. L. Jernigan, *Biochemistry*, in press.
42. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
43. S. R. Jordan and C. O. Pabo, *Science* **242**, 893 (1988).
44. R. M. Breyer, thesis, Massachusetts Institute of Technology, Cambridge (1988).
45. J.-L. Fauchere and V. Pliska, *Eur. J. Med. Chem.-Chim. Ther.* **18**, 369 (1983).

46. We thank C. O. Pabo and S. Jordan for coordinates of the NH_2-terminal domain of λ repressor and its operator complex. We also thank P. Schimmel for the use of his graphics system and J. Burnbaum and C. Francklyn for assistance. Supported in part by NIH grant AI-15706 and predoctoral grants from NSF (J.R.-O.) and Howard Hughes Medical Institute (W.A.L.).

rticle No. jmbi.1998.2362 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (1999) 286, 1771–1785

JMB

Effect of the Extra N-terminal Methionine Residue on the Stability and Folding of Recombinant α-Lactalbumin Expressed in *Escherichia coli*

Tapan K. Chaudhuri[1], Katsunori Horii[2], Takao Yoda[1], Munehito Arai[1], Shinji Nagata[3], Tomoki P. Terada[1], Hidefumi Uchiyama[4], Teikichi Ikura[1], Kouhei Tsumoto[2], Hiroshi Kataoka[3], Masaaki Matsushima[5], Kunihiro Kuwajima[1]* and Izumi Kumagai[2]

[1]*Department of Physics, Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan*

[2]*Department of Biomolecular Engineering, Tohoku University, Sendai 980-8579, Japan*

[3]*Department of Biotechnology, Faculty of Agriculture & Life Sciences, University of Tokyo, Tokyo, 113-8658, Japan*

[4]*Department of Biological Science and Technology, Science University of Tokyo, 2941 Yamazaki, Noda, Chiba 278-8510, Japan*

[5]*Rational Drug Design Laboratories, Fukushima 960-1242, Japan*

*Corresponding author

The structure, stability, and unfolding-refolding kinetics of *Escherichia coli*-expressed recombinant goat α-lactalbumin were studied by circular dichroism spectroscopy, X-ray crystallography, and stopped-flow measurements, and the results were compared with those of the authentic protein prepared from goat milk. The electric properties of the two proteins were also studied by gel electrophoresis and ion-exchange chromatography. Although the overall structures of the authentic and recombinant proteins are the same, the extra methionine residue at the N terminus of the recombinant protein remarkably affects the native-state stability and the electric properties. The native state of the recombinant protein was 3.5 kcal/mol less stable than the authentic protein, and the recombinant protein was more negatively charged than the authentic one. The recombinant protein unfolded 5.7 times faster than the authentic one, although there were no significant differences in the refolding rates of the two proteins. The destabilization of the recombinant protein can be fully interpreted in terms of the increased unfolding rate of the protein, indicating that the N-terminal region remains unorganized in the transition state of refolding, and hence is not involved in the folding initiation site of the protein. A comparison of the X-ray structures of recombinant α-lactalbumin determined here with that of the authentic protein shows that the structural differences between the proteins are confined to the N-terminal region. Theoretical considerations for the differences in the conformational and solvation free energies between the proteins show that the destabilization of the recombinant protein is primarily due to excess conformational entropy of the N-terminal methionine residue in the unfolded state, and also due to less exposure of hydrophobic surface on unfolding. The results suggest that when the N-terminal region of a protein has a rigid structure, expression of the protein by *E. coli*, which adds the extra methionine residue, destabilizes the native state through a conformational entropy effect. It also shows that differences in the electrostatic interactions of the N-terminal amino group with the side-chain atoms of Thr38, Asp37, and Asp83 bring about a difference in the pK_a value of the N-terminal amino group between the proteins, resulting in a greater negative net charge of the recombinant protein at neutral pH.

Keywords: recombinant goat α-lactalbumin; extra N-terminal methionine residue; protein folding; X-ray crystallographic study; conformational entropy

Abbreviations used: GdnHCl, guanidine hydrochloride; CD, circular dichroism; N, native; U, unfolded; SDS, sodium dodecyl sulfate; CNBr, cyanogen bromide; UV, ultraviolet; ASA, accessible surface area.

E-mail address of the corresponding author: kuwajima@phys.s.u-tokyo.ac.jp

Introduction

The N-terminal sequence of a recombinant protein expressed in *Escherichia coli* is known to start with formyl-methionine (Marcker & Sanger, 1964), which is in most cases subsequently processed by deformylase enzyme (Adams, 1968; Takeda & Webster, 1968), and removed by methionine aminopeptidase to finally produce the N-terminal methionine-free recombinant protein. However, removal of the N-terminal methionine does not always take place, and about half of *E. coli*-expressed proteins contain the extra N-terminal methionine residue, because the aminopeptidase action depends on the nature of the penultimate amino acid residue (Moerschell *et al.*, 1990). Therefore, the effect of the N-terminal methionine residue, when present, on the structure, stability and folding of *E. coli*-expressed recombinant proteins should be an important issue in biophysical and molecular biological studies that use such recombinant proteins, although this has not been taken seriously in most cases.

The biological and physicochemical properties of the methionylated proteins expressed in *E. coli* may differ from those of the authentic proteins that do not have the N-terminal methionine. For example, recombinant hen egg-white lysozyme contains the N-terminal methionine residue (Miki *et al.*, 1987; Mine *et al.*, 1997) and has lower solubility and stability than the authentic form (Imoto *et al.*, 1987). Similarly, recombinant apomyoglobin expressed in *E. coli* contains the extra N-terminal methionine residue and is less stable than the authentic protein (Hargrove *et al.*, 1994). On the other hand, the presence of the extra N-terminal methionine or the extension or truncation of the N-terminal residues does not interfere with the native-state stability in certain other globular proteins (Kordel *et al.*, 1989; Duverger *et al.*, 1991). In recombinant ribonuclease A, the extra N-terminal methionine is even known to stabilize the native structure (Schultz & Baldwin, 1992; Aronsson *et al.*, 1995). However, details of the effects of the extra N-terminal methionine residue on the structure, stability, and folding of the proteins have not yet been well understood.

α-Lactalbumin is a milk Ca^{2+}-binding protein, which consists of 123 amino acid residues and has a molecular mass of 14,200 Da. The three-dimensional structure of α-lactalbumin from several mammalian species, including goat, cow, guinea pig, and human, has been determined by X-ray crystallographic analysis (Acharya *et al.*, 1991; Pike *et al.*, 1996), and it is very similar to the structure of c-type lysozyme, a homologous protein. α-Lactalbumin has been used actively as a model protein in studies of protein folding (Sugai & Ikeguchi, 1994; Kuwajima, 1989, 1996; Vanderheeren & Hanssens, 1994; Uchiyama *et al.*, 1995; Schulman & Kim, 1996; Arai & Kuwajima, 1996; Schulman *et al.*, 1997; Wilson *et al.*, 1996; Shimizu *et al.*, 1996; Balbach *et al.*, 1996; Katsumata *et al.*, 1996; Kataoka *et al.*, 1997; Kuhlman *et al.*, 1997; Wu & Kim, 1997; Pfeil, 1998; Ikeguchi *et al.*, 1998), because this protein readily adopts a molten globule state, which is known to be identical with a folding intermediate (Kuwajima, 1989, 1996; Ptitsyn, 1995), under a variety of conditions, including those at a low pH, at a moderate concentration of guanidine hydrochloride (GdnHCl), and in the absence of Ca^{2+} and other salts (Kuwajima, 1989, 1996). Recombinant α-lactalbumin expressed in *E. coli*, though containing the extra N-terminal methionine, has often been used in these studies of protein folding. A recent study has, however, shown that like recombinant hen egg-white lysozyme, recombinant bovine α-lactalbumin is less stable than the authentic protein, although the lactose synthase regulatory activities of the recombinant and authentic proteins have been shown to be identical with each other (Ishikawa *et al.*, 1998).

Here, we show that *E. coli*-expressed recombinant goat α-lactalbumin is destabilized by the presence of the extra N-terminal methionine residue by as much as 3.5 kcal/mol and has a more negative electric net charge than the authentic protein. It is concluded that the destabilization of the recombinant protein is primarily brought about by an extra conformational entropy of the methionyl residue in the unfolded state and that the more negative charge of the recombinant protein is caused by a decrease in the pK_a value of the N-terminal amino group. Because the N-terminal methionine remarkably destabilizes recombinant α-lactalbumin, the role of the N terminus in the folding of this protein has also been investigated by stopped-flow circular dichroism (CD) studies of the unfolding and refolding kinetics of the recombinant and authentic proteins. The destabilization of the recombinant protein is shown to be entirely interpreted in terms of an increase in the unfolding rate, indicating that the N terminus is not involved in the folding initiation site of α-lactalbumin.

Results

Structure of folded recombinant goat α-lactalbumin

The recombinant wild-type protein was expressed in *E. coli* as inclusion bodies with a high yield (15 mg per litre of culture). The protein was solubilized in 8 M urea and refolded in a redox buffer in the absence of urea at pH 8.5 and 4 °C. The process of refolding was monitored by reversed-phase HPLC (Uchiyama *et al.*, 1995), and the folded protein was purified (see Materials and Methods). The peptide and aromatic CD spectra of the recombinant protein were measured under native conditions (0 M GdnHCl (pH 8.0) at 25 °C), and compared with those of authentic goat α-lactalbumin (Figure 1(a) and (b)). There is no significant difference in the CD spectra between the proteins in the aromatic and peptide regions, so that the secondary and tertiary structures of the

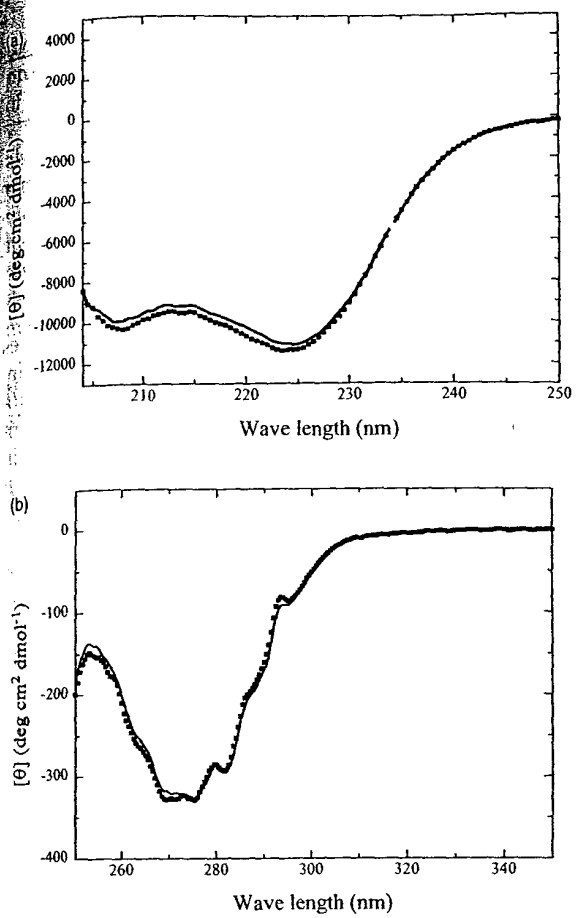


Figure 1. (a) Far and (b) near-UV CD spectra of authentic and recombinant goat α-lactalbumin measured in the presence of 1 mM $CaCl_2$ at pH 7.0 and 25 °C. The continuous line denotes the authentic protein and the filled squares denote the recombinant protein.

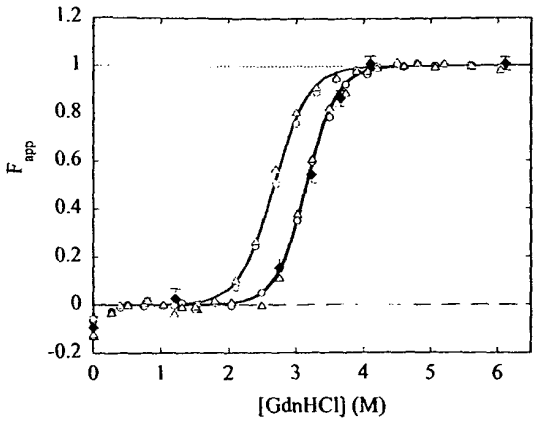


Figure 2. GdnHCl-induced unfolding transition curves for authentic, recombinant, and methionine-free goat α-lactalbumin. The unfolding was carried out at 25 °C in the presence of 1 mM $CaCl_2$, 50 mM NaCl, and 50 mM sodium cacodylate (pH 7.0), and the transitions were monitored by both far and near-UV CD measurements. Apparent fractions of unfolded species (F_{app}) were plotted against the concentration of GdnHCl. Open black circles and open red cirles denote the F_{app} values measured at 222 nm for authentic and recombinant proteins, respectively; open black triangles and open red triangles represent the F_{app} values measured at 270 nm for authentic and recombinant proteins, respectively; and the F_{app} values of the methionine-free recombinant protein measured at 222 nm are presented by filled blue diamonds.

two proteins are essentially identical with each other. This conclusion is confirmed by the X-ray crystallographic structure of recombinant goat α-lactalbumin (see below). The results thus indicate that the folded recombinant protein is correctly folded into the native structure. A study has also shown that the lactose synthase regulatory activity of the folded recombinant protein is the same as that of authentic α-lactalbumin (Uchiyama *et al.*, 1995).

Equilibrium unfolding

The GdnHCl-induced equilibrium unfolding transition of the folded recombinant protein was studied by the peptide and aromatic CD spectra, and the results were compared with those of authentic goat α-lactalbumin. Figure 2 shows the unfolding transition curves of the two proteins measured by the CD ellipticities, at 222 and 270 nm, and these ellipticities are expressed by the apparent fractional extent (F_{app}) of unfolding as a

From Figure 2, the unfolding transition curves measured at 222 and 270 nm are coincident with each other in authentic and recombinant α-lactalbumin, indicating that the unfolding transitions of the two proteins are well represented by a two-state mechanism, in which only the native (N) and the fully unfolded (U) states are populated in the transition zone as:

$$N \overset{K_U}{\rightleftharpoons} U \qquad (1)$$

Here K_U is the equilibrium constant of unfolding and relates to the free energy change, ΔG_U, of the unfolding transition as:

$$K_U = \exp(-\Delta G_U/RT) \qquad (2)$$

where R and T are the gas constant and the absolute temperature, respectively, and ΔG_U is assumed to be linearly dependent on GdnHCl concentration (C) as:

$$\Delta G_U = \Delta G_U^{H_2O} - mC = m(C_m - C) \qquad (3)$$

where $\Delta G_U^{H_2O}$ is the ΔG_U in the absence of the denaturant, C_m is the C at the midpoint of the unfolding transition, and m represents the dependence of ΔG_U on C and is a measure of the cooperativity of the transition (Pace, 1986). From

Table 1. Equilibrium u .g transition parameters of goat α-lactalbum

Name of protein	$\Delta G_U^{H_2O}$ (kcal/mol)	m (kcal/ mol M)	C_m (M)	$\Delta\Delta G_U^{H_2O}$ (kcal/mol)	$\Delta\Delta G_U$ (kcal/mol) at 3.2 M GdnHCl
Authentic goat α-lactalbumin	13.8 ± 0.7	4.4 ± 0.2	3.15 ± 0.01	-	-
Recombinant goat α-lactalbumin	10.4 ± 0.5	3.9 ± 0.2	2.67 ± 0.01	−3.5	−1.9

by F_{app} is given as a function of C as:

$$F_{app}(C) = \frac{\exp[-m(C_m - C)/RT]}{1 + \exp[-m(C_m - C)/RT]} \quad (4)$$

The values of m, C_m, and hence $\Delta G_U^{H_2O}$, for recombinant and authentic α-lactalbumin were calculated from the data of Figure 2 by the non-linear least-squares method. The unfolding parameters m, C_m, and hence $\Delta G_U^{H_2O}$, thus obtained are summarized in Table 1. The continuous lines in Figure 2 are the curves theoretically drawn with the parameter values of Table 1, and show excellent agreement between theory and the experimental data.

Figure 2 also shows that the unfolding transition of the recombinant protein occurs at a remarkably lower concentration of GdnHCl ($C_m = 2.7$ M) than the transition of authentic α-lactalbumin ($C_m = 3.2$ M). The difference in ΔG_U ($\Delta\Delta G_U$) is −3.5 kcal/mol at 0 M GdnHCl and −1.9 kcal/mol at 3.2 M GdnHCl, which is the C_m for the authentic protein (Table 1). Therefore, the folded recombinant protein is remarkably less stable than authentic α-lactalbumin, although their native structures are practically identical as evidenced by the CD spectra and X-ray structural analysis.

Gel electrophoresis and ion-exchange chromatography

In order to investigate further differences between recombinant and authentic α-lactalbumin, the electrophoretic and ion-exchange chromatographic behavior of the two proteins were investigated. Figure 3(a) shows electrophoretic patterns in a non- denaturing polyacrylamide gel at pH 9.4. It can be seen that the electrophoretic mobility of the recombinant protein is significantly greater than that of the authentic protein. Figure 3(b) shows the elution profiles of recombinant and authentic α-lactalbumin in an anion-exchange HPLC using a RESOURCE™ Q column (Pharmacia Biotech) with a linear gradient from 0 M to 0.5 M NaCl in the presence of 10 mM NaH_2PO_4-Na_2HPO_4 buffer (pH 7.0). The retention time is longer for the recombinant protein (22.9 minutes) than for the authentic one (19.6 minutes). Both of these results indicate that the recombinant protein is more negatively charged. These differences in the electric properties of the two proteins, however, disappear in the U state in 8 M urea. The electrophoretic mobilities and the chromatographic retention times of the proteins were found to be identical in the presence of 8 M urea (data not shown). Therefore,

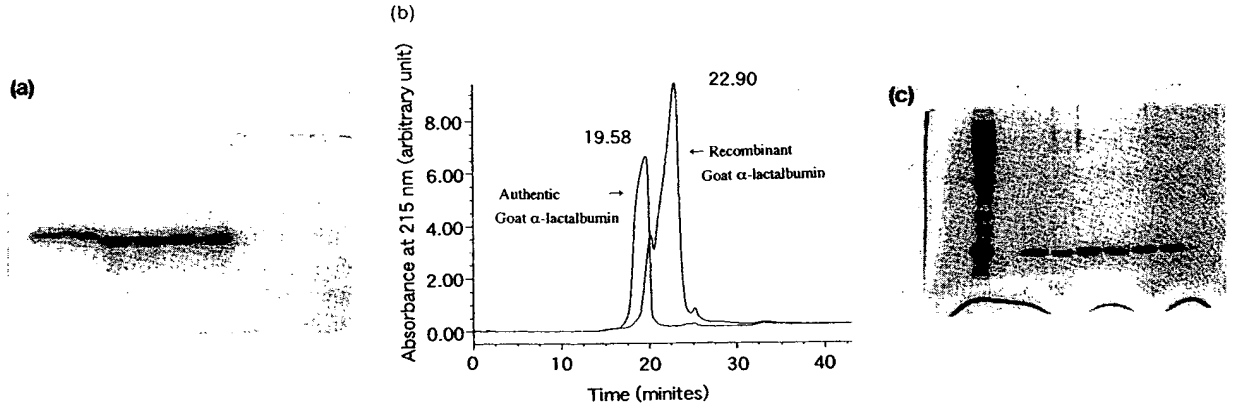


Figure 3. (a) Non-denaturing polyacrylamide gel electrophoresis of authentic and recombinant goat α-lactalbumin. The electrophoresis was carried out at pH 8.9 using 12% (w/v) acrylamide gel at room temperature. About 5 μg protein samples were applied in each lane. Lanes 1 and 2 contain authentic protein, lanes 3 and 4 contain recombinant protein, and lanes 5 and 6 contain an equimolecular mixture of authentic and recombinant protein. Lanes are numbered from left to right in (a) and (c). (b) Superimposed HPLC pattern of authentic and recombinant goat α-lactalbumin. The HPLC was performed with a Resource-Q anion exchange column at pH 7.0 using a linear gradient of 0 M-0.5 M NaCl containing 10 mM NaH_2PO_4-Na_2HPO_4 buffer. A 50 μl protein sample containing 50 and 60 μg of native and recombinant protein, respectively, was applied in the HPLC column. (c) SDS/polyacrylamide gel elecrophoresis of authentic and recombinant goat α-lactalbumin. The electrophoresis was carried out using 15% (w/v) acrylamide gel at room temperature. Approximately 10 μg of protein was applied in each lane. Lane 1 contains low molecular mass marker proteins, lane 2 is blank, lanes 3 and 4 contain the authentic protein, lanes 5 and 6 contain the recombinant protein and lanes 7 and 8 contain an equimolecular mixture of the authentic and recombinant proteins.

the difference in the electric charge between the proteins must be caused by the structural folding of the proteins into the native structure.

SDS/polyacrylamide gel electrophoresis was also carried out for the recombinant and authentic proteins using 15% acrylamide in the resolving gel (Figure 3(c)). The electrophoretic mobilities of the two proteins are the same within the experimental error, indicating that there is no significant difference in the molecular mass between the proteins.

N-terminal sequence and mass spectrometric analyses

In order to identify any differences in the amino acid sequence, we performed N-terminal sequencing and mass spectrometric analysis of the recombinant and authentic proteins. The N-terminal sequences of the first five residues of the two proteins have shown that recombinant α-lactalbumin contains an additional methionine residue. The results of the mass spectrometric analysis indicate that the difference in mass between the recombinant and authentic proteins is 133 (Figure 4), which is nearly equal to the mass of a single methionine residue (131.19), confirming the presence of the extra methionine residue in the recombinant protein. Therefore, the only chemical difference that brings about the difference in the electric charge between the two proteins in the N state is the presence or absence of the extra methionine residue at the N terminus, and this difference may also lead to the remarkable difference in stability between the proteins.

Methionine-free recombinant α-lactalbumin

In order to directly investigate the effect of the extra methionine residue on the electric properties and stability of the recombinant protein, methionine-free recombinant α-lactalbumin was prepared by cyanogen bromide (CNBr) cleavage. Because there is no methionine residue in authentic goat α-lactalbumin, only the extra N-terminal methionine of the recombinant protein is expected to be removed by the CNBr cleavage. The removal of the methionine was confirmed by N-terminal sequencing and mass spectrometric analysis (data not shown). The absence of other cleavage products was confirmed by SDS/polyacrylamide gel electrophoresis. The near and far-UV CD spectra of the methionine-free recombinant protein overlap with those of the authentic and original recombinant proteins (data not shown). The electrophoretic mobility in the native gel and the retention time for the anion-exchange chromatography were found to be identical with those of the authentic protein (data not shown). The stability of the methionine-free recombinant protein against the GdnHCl-induced unfolding was investigated, and the equilibrium unfolding transition of the methionine-free protein is shown in Figure 2. The unfolding transition curve coincides well with that of the authentic protein, and gives the same C_m and $\Delta G_U^{H_2O}$ values. As a control, the authentic protein was also subjected to the conditions of CNBr cleavage, and it was confirmed that the unfolding transition of the protein was not affected by the cleavage conditions (data not shown). These results thus clearly indicate that the observed destabilization and the difference in the electric charge of the recombinant protein is solely due to the presence of the extra N-terminal methionine residue.

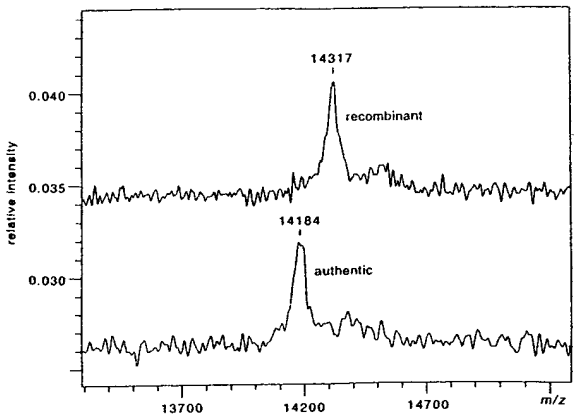


Figure 4. MALDI-TOF-MS mass spectroscopic pattern of authentic and recombinant goat α-lactalbumin. The upper trace is for recombinant and the lower one is for authentic protein.

Kinetics of refolding and unfolding

The above results indicate that the presence of the extra methionine residue at the N terminus of the recombinant protein decreases the relative stability of the N state by as much as 3.5 kcal/mol. Thus, it appears that both recombinant and authentic goat α-lactalbumin are useful for investigating the role of the N-terminal residue in the kinetic folding of α-lactalbumin. The kinetic unfolding and refolding reactions of the recombinant and authentic proteins were investigated by stopped-flow CD measurements. The unfolding and refolding reactions were induced by concentration jumps of GdnHCl from 1.0 to 5.4 M and from 5.5 to 0.5 M, respectively. The reactions were monitored by the ellipticity change at 225 nm at pH 7.0 and 25 °C. The kinetic progress curves for unfolding and refolding are shown in Figure 5(a) and (b), respectively, and the data were fitted by the non-linear least-squares method with the equation:

$$A(t) = A(\infty) + \Delta A_{obs} \sum \alpha_i \exp(-k_i t) \qquad (5)$$

where $A(t)$ and $A(\infty)$ are the observed values of the ellipticity at time t and infinite time, respectively, ΔA_{obs} is the observed total amplitude $[A(0) - A(\infty)]$, and k_i and α_i are the apparent first-

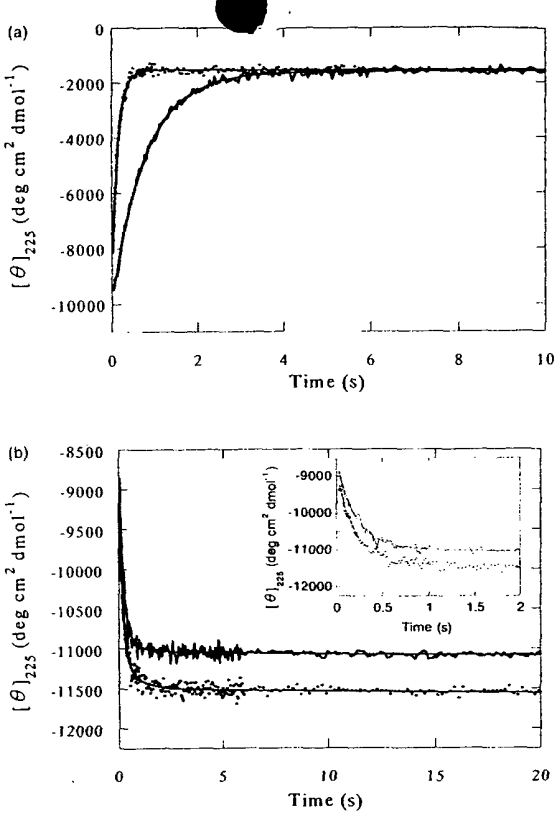


Figure 5. GdnHCl-induced (a) unfolding and (b) refolding kinetic progress curves of authentic and recombinant goat α-lactalbumin. The unfolding was initiated by a concentration jump from 1.0 M to 5.4 M and the refolding process was initiated by a concentration jump of 5.5 M to 0.5 M at 25°C in the presence of 1 mM $CaCl_2$, 50 mM NaCl, and 50 mM sodium cacodylate, pH 7.0, and the refolding and unfolding kinetics were monitored by the measurement of CD ellipticity at 225 nm using stopped-flow CD. The continuous line denotes authentic protein and the filled squares denote recombinant protein. (b) The inset shows the refolding progress curve within two seconds and the same notations are used for the transition curves. Theoretical kinetic progress curves are also shown in (a) and (b).

order rate constant and fractional amplitude, respectively, of the *i*th kinetic phase.

The kinetic progress curves for unfolding for both the recombinant and authentic proteins were well fitted to a single-exponential equation, and the apparent rate constants and the amplitudes for the two proteins are presented in Table 2. The kinetic progress curves for refolding were well fitted to the three-exponential equation, and the rate constants and the amplitudes are presented in Table 3. The unfolding reaction of recombinant α-lactalbumin is 5.7-times faster than that of the authentic protein, while there are no significant differences in the rate constants for the triphasic refolding reactions of the two proteins. Thus, it appears that the N-terminal end of goat α-lactalbumin is not essential for the kinetic folding of this protein (see Discussion).

Table 2. unfolding parameters of goat α-lactalbumin

Name of protein	k_1 (s^{-1})	$\alpha_1\Delta A_{obs}$ (deg cm^2 dmol^{-1})
Authentic goat α-lactalbumin	1.26 ± 0.01	−8384
Recombinant goat α-lactalbumin	7.18 ± 0.08	−8056

X-ray crystallographic study

In order to further investigate the differences in the folded structure between recombinant and authentic goat α-lactalbumin, an X-ray crystallographic analysis of the recombinant protein was performed, and the structure was compared with that reported for the authentic protein structure. The crystallographic data are summarized in Table 4. The space group of the crystal of the recombinant protein was altered to $P2_12_12$ from $P2_1$ in which the authentic protein was packed (Pike *et al.*, 1996). The number of protein molecules in the asymmetric unit was one, although there were two (Mol A and Mol B) in the authentic protein crystal. The final *R* and free *R* factors were 0.191 and 0.278 in the resolution range of 8.0 to 2.0 Å. The overall error was estimated at 0.19 Å by a Luzzati plot (Luzzati, 1952). As the space group is altered in the recombinant protein crystal, the N-terminal methionine may affect the molecular packing in the crystal. However, the interactions between the two independent authentic molecules (Mol A and Mol B) were found to be very similar to the interactions between the symmetry-related recombinant molecules (Figure 6).

The structural differences between the recombinant and the authentic proteins are shown in Figure 7, which represents the distances between the C^α atoms of the two molecules. The root-mean-square deviations of the main-chain atoms are 0.55 Å between the recombinant protein molecule

Table 3. Kinetic refolding parameters of goat α-lactalbumin

Name of protein	k_1 (s^{-1})	$\alpha_1\Delta A_{obs}$ (deg cm^2 dmol^{-1})	k_2 (s^{-1})	$\alpha_2\Delta A_{obs}$ (deg cm^2 dmol^{-1})	k_3 (s^{-1})	$\alpha_3\Delta A_{obs}$ (deg cm^2 dmol^{-1})
Authentic goat α-lactalbumin	0.11 ± 0.05	64.28	1.3 ± 1.1	145	4.9 ± 0.3	2282
Recombinant goat α-lactalbumin	0.09 ± 0.04	70.8	1.3 ± 0.4	335	5.7 ± 0.4	2234

Table 4. Crystallization, data collection, and refinement statistics of recombinant goat α-lactalbumin

A. Crystallization	
Reservoir solution	1.0 mM $CaCl_2$ 16-20% PEG8000 0.05 M KH_2PO_4 pH 6.0
Protein concentration	20 mg protein/ml
Temperature (°C)	20
B. Crystal data	
Space group	$P2_12_12$
a, b, c (Å)	44.9, 88.9, 32.2
In an asymmetric unit	1
X-ray generator	Cu target (4.5 kW)
Resolution at measurements (Å)	1.75
Total number of ind. refl.	12,533
R_{merge}[a]	0.069
Completeness (%)	92.5
C. Structure determination	
Method	Mol. replacement
Model structure	Baboon α-LA
Software	X-PLOR 3.1
D. Refinement	
Software	X-PLOR 3.1
Resolution range (Å)	8.0-2.0
R-factor[b]	0.191
R_{free}[c]	0.278
Rms deviations in:	
Bond length (Å)	0.010
Bond angles (°)	1.554

[a] $R_{merge} = \Sigma_h \Sigma_i |I(h, i) - \langle I(h)\rangle| / \Sigma_h \Sigma_i I(h, i)$, where $I(h, i)$ is the intensity value of the ith measurement of h and $\langle I(h)\rangle$ is the corresponding mean value of $I(h)$ for all i measurements.

[b] R-factor $= \Sigma||F_{obs}| - |F_{calc}|| / |F_{obs}|$, where $|F_{obs}|$ and $|F_{calc}|$ are observed and calculated structure factor amplitude respectively.

[c] R_{free} is the same as R-factor, but for a 10% subset of all reflections.

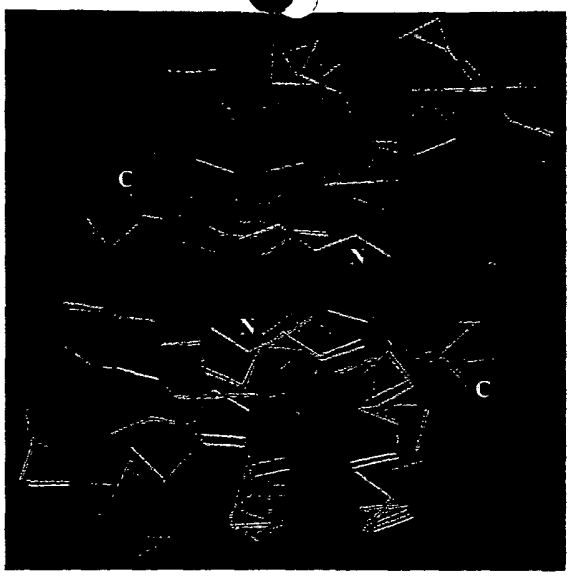


Figure 6. The molecular packings of recombinant and authentic goat α-lactalbumin in the crystals. The main-chain atoms of Mol A (orange) were superimposed on those of the recombinant protein molecule (blue). The same transformation matrixes were applied on Mol B (yellow). The Figure shows two of the symmetry-related recombinant protein molecules, and Mol A and B are overlaid. The space group was $P2_12_12$ in the recombinant protein crystal and $P2_1$ in the authentic protein crystal. The interactions in both crystals were very similar.

and Mol A, and 0.63 Å between the recombinant molecule and Mol B. These values are larger than the root-mean-square deviation between Mol A and Mol B (0.27 Å). From Figure 7, we can see that the intermolecular interactions remarkably affect the structure of the N-terminal and loop regions of the protein, especially between residues 105 and 110, but that the overall structures of the recombinant and authentic proteins are essentially identical, supporting previous observations of the same CD spectra of the proteins in solution.

The structures around the N termini of the recombinant protein and the two molecules of the authentic protein are shown in Figure 8, and we may see structural differences that give rise to the differences in the electric properties and stability between them. The N-terminal amino group strongly interacts with the side-chain atoms of Thr38 and Gln39 in Mol A (Figure 8(a)) and Thr38 in Mol B through hydrogen bonds and/or salt bridges (Figure 8(b)). A similar interaction can also be observed in the recombinant protein, in which the N-terminal amino group is bound to the side-chain of Gln39 by a hydrogen bond (Figure 8(c)), but this interaction may be significantly stronger than the corresponding interaction in the authentic protein (see Discussion). It can also be seen from Figure 8(c) that the methionine side-chain of the recombinant protein is directly in contact with the side-chain of Gln2, and that the orientation of the methionine side-chain is fixed by the hydrogen bonds between the N-terminal amino group and the side-chain of Gln39, and between the main-

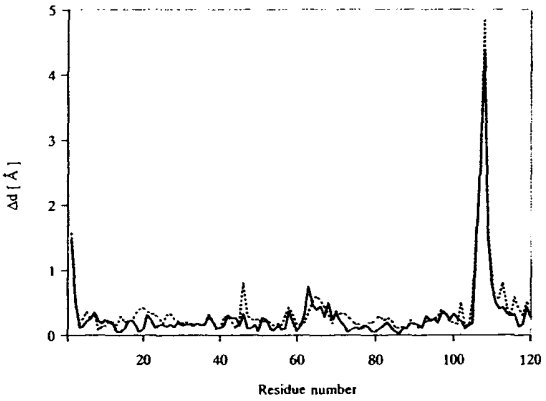


Figure 7. The structural differences between the corresponding C^{α} atoms of the recombinant and authentic protein molecules. Differences were observed in the N-terminal residues and the flexible loop residues of 105-110. The loop residues of the recombinant protein were affected by the neighboring molecules in the crystal.

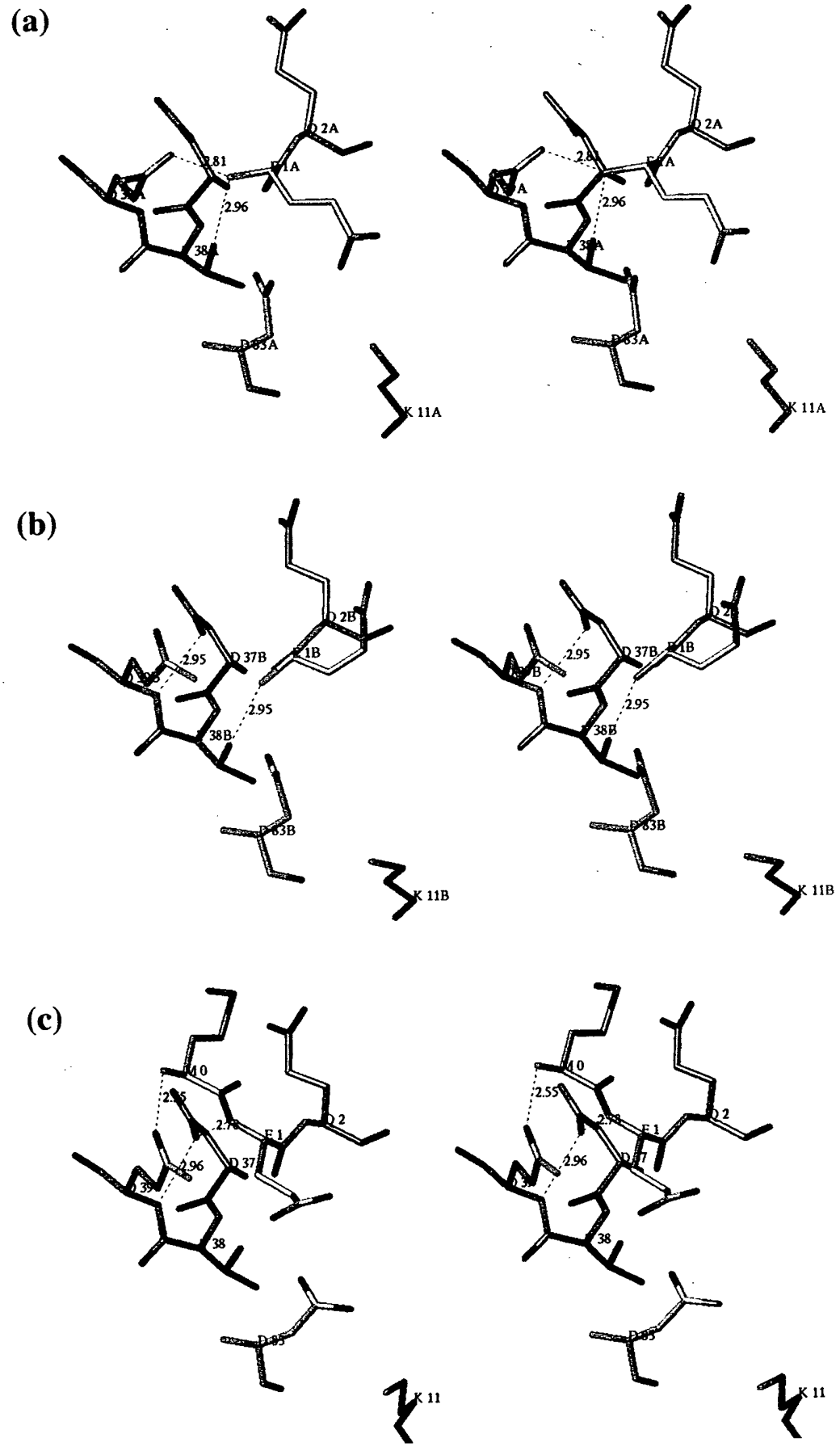


Figure 8 (*legend opposite*)

chain amido group of Glu1 and the carboxyl group of Asp37. The side-chains of the methionine and Gln2 residues weakly interact with a neighboring protein molecule by van der Waals contacts (Figure 8(c)). The main-chain conformations of residues Glu1 and Gln2 of the recombinant protein are almost the same as those of the authentic molecules. It is interesting that very similar intermolecular interactions are found in the recombinant and authentic protein molecules, although their crystallographic packings are different. The side-chain of Glu1 is folded into the inside of the recombinant molecule and interacts with the amino group of Lys11, and this conformation is similar to that in Mol A, but the Glu1 side-chain is shifted further from the amino group of Lys11 in the recombinant protein. The corresponding conformation of the Glu1 side-chain of Mol B is affected by the positively charged His107 side-chains of the symmetry-related molecules (Mol A and Mol B) in the crystal of the authentic protein. Without the presence of these positive charges near Glu1 of Mol B, the conformation of the side-chain is similar to that of the recombinant molecule and Mol A of the authentic protein.

Discussion

The present results show that the presence of the additional N-terminal methionine residue in recombinant α-lactalbumin expressed in *E. coli* remarkably decreases the stability of the native protein and increases the apparent net negative charge in the native state. Many proteins expressed in *E. coli* have the N-terminal methionine residue, although whether or not the methionine residue is present depends on the next residue in the recombinant protein (Miller *et al.*, 1987). Thus, the effect of the N-terminal methionine on the structure, stability, and other properties of an *E. coli*-expressed protein is important when we use the recombinant protein in biophysical and molecular biological studies. Such effects of the N-terminal methionine residue have, however, been ignored so far in most cases. As far as we are aware, the present study is the first of its kind in which the effect of the N-terminal methionine residue is thoroughly investigated by equilibrium unfolding and kinetic unfolding-refolding studies as well as by the CD and X-ray crystallographic analyses. Only recently, Ishikawa *et al.* (1998) have reported the effect of the N-terminal methionine on the thermal unfolding of bovine α-lactalbumin, and their results and the present results should be complementary to each other.

Because the N-terminal methionine affects the native state stability of α-lactalbumin, the effects of the methionine on the kinetics of refolding and unfolding of the protein were investigated. The results of our study show that the recombinant protein unfolds 5.7 times faster than the authentic one, whereas the rates of refolding remain the same. The results should provide an insight into the role of the N terminus in the folding mechanism for α-lactalbumin.

We will first consider here the structural aspects that explain differences in the native state stability and the electric properties between the recombinant and authentic proteins on the basis of our CD and X-ray crystallographic data. We will then discuss the mechanism of folding for goat α-lactalbumin on the basis of the kinetic unfolding and refolding data.

Stabilities of recombinant and authentic α-lactalbumin

Structure around the N terminus

The results show that although recombinant and authentic α-lactalbumin follow the two-state unfolding transition (Figure 2), the recombinant protein is less stable than the authentic one by as much as 3.5 kcal/mol in the absence of GdnHCl. In order to understand this stability difference ($\Delta\Delta G_U$), we determined the X-ray structure of the recombinant protein, and this structure was compared with the X-ray structure of the authentic protein determined by Pike *et al.* (1996). The overall structures of the two proteins are essentially identical with each other, being consistent with the identical CD spectra of the proteins, and the structural differences between the proteins have been found to be localized in the N-terminal and the 105-110 loop regions (Figure 7). Because the structural differences in the 105-110 loop region, which is very flexible in the N state, are likely to be caused by a difference in the crystallographic packing between the proteins (Acharya *et al.*, 1991; Harata & Muraki, 1992; Pike *et al.*, 1996), we have concentrated our attention on the structural differences in the N-terminal region and investigated any interactions that are present in the authentic protein but missing in the recombinant one. Our data, however, show that there are no such interactions identified in the X-ray structures. From Figure 8, it can be seen that the N-terminal amino group of

Figure 8. Stereo views of the N-terminal region of authentic goat α-lactalbumin, (a) Mol A and (b) Mol B, and (c) recombinant goat α-lactalbumin. The main-chain structures of these three molecules were very similar. But the mainchain of the recombinant protein was shifted to the outside of the molecule, compared with those of the authentic molecules. The side-chain conformation of Gln2 of Mol B is different from those of the others. The side-chain of Asp83 of the recombinant protein was not clearly seen in the electron density map, and the *B*-factors of the side-chain atoms were high. Therefore, the model coordinates could not be explicitly determined. Certain distances are shown in Å, and the residues are shown by the one-letter code.

Glu1 of the authentic protein is hydrogen-bonded with two side-chain oxygen atoms of Thr38 and Gln39 in Mol A and with a side-chain oxygen atom of Thr38 in Mol B. A similar hydrogen bond is also observed in the recombinant protein between the N-terminal amino group and Gln39, and the length of the hydrogen bond is smaller than that in the authentic protein, suggesting that the hydrogen bond is even stronger in the recombinant protein (Figure 8(c)). Although the hydrogen bond between the N-terminal amino group and Thr38 is missing in the recombinant protein, there is an alternative hydrogen bond between the main-chain amido group of Glu1 and the carboxyl group of Asp37. The degrees of the packing interactions of the side-chain atoms are also very similar in the N-terminal regions of the two proteins. The side-chains are closely packed in both proteins. Furthermore, contributions of electrostatic interactions around the N termini to the destabilization of the recombinant protein will be shown to be negligibly small, although they are related to the difference in the electric net charge between the proteins (see below). Therefore, the observed destabilization cannot be interpreted in terms of the presumed interactions missing in the native structure of the recombinant protein.

Conformational entropy of the methionine residue and solvation free energies

If the destabilization of the recombinant protein cannot be simply explained by the interactions identified in the X-ray structures of the recombinant and authentic proteins, what makes the recombinant protein less stable? At this point, it should be noted that the N-terminal residues of both the recombinant and authentic proteins are involved in a rigid structure, so that all the atoms of the residues can be traced in the electron density maps of the proteins by X-ray crystallographic analysis. The *B*-factors of the backbone atoms of the N-terminal methionine residue of the recombinant protein were found to range from 31 to 35 Å^2. The values are much larger than those of the residues buried inside the protein molecule (8-15 Å^2), but are smaller than those of the fully exposed residues in flexible loop regions. This means that the presence of the additional methionine residue in the recombinant protein destabilizes the native state through an entropic effect, which arises from an additional conformational entropy of the methionine residue in the U state. Because the structure around the N terminus is rigid in the N state of the recombinant protein, the additional methionine residue leads to an increase in entropy on unfolding. Thus, the free energy change of unfolding (ΔG_U), which is the difference in the free energy between the N and U states, decreases, and hence the N state of the recombinant protein is destabilized.

estimated at 20 cal/(mol K) by Oobatake & Ooi (1993) from an analysis of hydration and heat stability effects on the unfolding of 14 globular proteins, and this corresponds with the free energy change of −5.9 kcal/mol at 25 °C. This value is close to but lower than the observed difference ($\Delta\Delta G_U = -3.5$ kcal/mol) in ΔG_U between the recombinant and authentic proteins. We have, however, ignored the contribution of the hydration free energy, ΔG_h^u, and the enthalpic contribution of the conformational unfolding, ΔH_c^u, which mostly arises from the van der Waals interaction energy, to the $\Delta\Delta G_U$ (Oobatake & Ooi, 1993). These contributions are expected to be proportional to the change in the accessible surface area of the methionine residue on unfolding (Oobatake & Ooi, 1993) and may explain the above difference between the expected contribution of the conformational entropy ($-T\Delta S_c^u$) and the observed $\Delta\Delta G_U$. The values of ΔG_h^u and ΔH_c^u of the N-terminal methionine residue were calculated by the method described by Oobatake & Ooi (1993), and they were −1.2 and 3.3 kcal/mol for ΔG_h^u, and ΔH_c^u, respectively, so that the free energy change of unfolding of the methionine residue (ΔG^u) was estimated at −3.8 kcal/mol (see equation (8)), which was in good agreement with the observed $\Delta\Delta G_U$ (see Materials and Methods). The contribution of other residues to the $\Delta\Delta G_U$ was also estimated, and it was less than 1 kcal/mol (see Materials and Methods), confirming that the increase in the conformational entropy of the N-terminal methionine residue on unfolding is a dominant factor determining the $\Delta\Delta G_U$.

In the above argument of $\Delta\Delta G_U$, however, we have implicitly assumed that the U state is fully unfolded in both the recombinant and authentic proteins. Thus, if there is a difference in the U-state structure between the proteins, such a difference may also contribute to the $\Delta\Delta G_U$. In fact, the *m* value of the equilibrium unfolding transition is found to be smaller for the recombinant protein (Table 1). Lower values of *m* are usually thought to be due to less exposure of hydrophobic surface on unfolding. Because the native structure is essentially identical between recombinant and authentic α-lactalbumin, the less exposure of hydrophobic surface must be due to a difference in the U-state structure, and the U state of the recombinant protein less exposes the hydrophobic surface than that of the authentic one. Similar effects of hydrophobic replacements of amino acid residues on the U-state structure have also been reported in staphylococcal nuclease (Shortle, 1996). The less difference in solvent exposed hydrophobic surface means a smaller difference in ΔG_U. Therefore, this may also be a factor determining the $\Delta\Delta G_U$ between the recombinant and authentic proteins.

Comparison with other proteins

authentic bovine α-lactalbumin using thermal denaturation measurements of the proteins. They have shown that the destabilization of the recombinant protein is caused by an entropic effect because the enthalpy change of the thermal unfolding is the same for the two proteins, and their result is fully consistent with our proposal regarding the destabilization of the recombinant protein described above. Although Ishikawa *et al.* (1998) have attributed the destabilization of the recombinant protein to a weakening of the apparent Ca^{2+}-binding strength, this interpretation seems to be nothing more than a rephrasing of the destabilization of the protein because the apparent Ca^{2+}-binding strength of α-lactalbumin is known to be linked to the N ⇌ U equilibrium of the apo protein (Hiraoka & Sugai, 1985). Our X-ray structural data show that there is no essential difference in the structure of the Ca^{2+}-binding site between the authentic and recombinant proteins, indicating that the weakening of the apparent Ca^{2+}-binding strength of the recombinant protein is caused by a destabilization of its apo form.

There have been several other reports of the effect of additional residues at the N terminus on the native-state stability of recombinant proteins, and a comparison of these with the present results will provide insight into a rule relating to the effects of an extra methionine residue in the proteins. Hargrove *et al.* (1994) have observed that the recombinant apomyoglobin expressed in *E. coli* is less stable than the authentic protein. They have also shown that the N terminus of recombinant apomyoglobin contains an extra methionine residue and that the structure around the N terminus is rigid. Polyhistidine tags in the N and C-terminal regions of Arc repressor (Milla *et al.*, 1993, 1995) have little effect on the stability and folding of the protein, whereas the polyhistidine tags of CspA alter the folding behavior by interacting with the wild-type portion of the protein (Reid *et al.*, 1998). The X-ray crystallographic structures of the Arc repressor (Raumann *et al.*, 1994) and CspA (Goldstein *et al.*, 1990) have shown that the structure around the N-terminal residue in CspA is rigid, whereas that of Arc repressor is flexible. The N-terminal region of staphylococcal nuclease is flexible (Hynes & Fox, 1991), and it has been reported that a 19-residue pro-peptide in the N-terminal region of the nuclease does not significantly destabilize the N state of the recombinant protein (pro staphylococcal nuclease; Suciu & Inouye, 1996). Therefore, these studies together with the our study strongly suggest that when the structure around the N-terminal residue of a protein is rigid, the addition of extra residues at the N terminus destabilizes the N state of the protein. On the other hand, when the structure is flexible, the extra residues do not interfere with the native-state stability. From these experimental results, we can thus conclude that when the N-terminal region of residue, destabilizes the N state, but that when the N-terminal region is flexible, expression of the protein by *E. coli* does not interfere with the native-state stability.

Electric properties of authentic and recombinant α-lactalbumin

The results of the electrophoresis and ion-exchange chromatography show that the recombinant protein is more negatively charged than the authentic one. It is understood, however, that the side-chain of a methionine residue does not ionize at neutral pH, so that there is no difference in the number of ionizable groups between the authentic and recombinant proteins. In fact, our electrophoresis and ion-exchange chromatography data show that there is no difference in the electric charge between the proteins in the presence of 8 M urea. This means that some of the ionizable groups that have a pK_a near 7.0 experience a change in pK_a due to the structural folding of the protein. There are two such ionizable groups, the imidazole group of histidine and the N-terminal amino group, which have intrinsic pK_a values of 6.5 and 8.0, respectively. If we compare the X-ray structures of the two proteins, no significant differences are observed near the histidine side-chains. However, there is a noticeable difference in the structures around the N-terminal amino groups. The N-terminal amino group is hydrogen-bonded to the oxygen atom of Thr38 and is closer to the side-chains of Asp37 and Asp83 in the authentic protein (Figure 8), and both of these may increase the pK_a value of the N-terminal amino group through electrostatic interactions. A study of the pH-dependence of the unfolding transition of authentic bovine α-lactalbumin has shown that the N-terminal amino group of the protein has an abnormally high pK_a value ($pK_a = 8.9$) in the N state, which is normalized on unfolding from the N to the molten globule state (Kuwajima *et al.*, 1981).

It should also be mentioned that the ΔpK_a of the N-terminal group between the recombinant and authentic proteins leads to a difference in the native-state stability between the proteins, but this stability difference is expected to be much smaller than the $\Delta\Delta G_U$ estimated from equation (6) at pH 7.0. The stability difference ($\Delta\Delta G_U(\Delta pK_a)$) due to the ΔpK_a is known to be given by:

$$\Delta\Delta G_U(\Delta pK_a) = RT \ln[(1 - K_a(\text{rec})/[H^+])/(1 - K_a(\text{auth})/[H^+])] \quad (6)$$

where K_a(rec) and K_a(auth) are the dissociation constants of the N-terminal amino groups of the recombinant and authentic proteins, respectively, and $[H^+]$ is the hydrogen-ion concentration (Tanford, 1970). If we assume that the pK_a(rec) and pK_a(auth) are 8.0 and 8.9, respectively, the above equation gives a $\Delta\Delta G_U(\Delta pK_a)$ of 0.06 kcal/mol at

N-terminal amino group reasonably interprets the differences in the electric properties between the proteins observed by electrophoresis and ion-exchange chromatography, but it is not sufficient for interpreting the stability difference between the proteins.

Folding of goat α-lactalbumin

Because the presence of the N-terminal methionine residue in the recombinant protein changes the thermodynamic stability of the native state, this system is useful for investigating the role of the N terminus in the folding of α-lactalbumin. We thus investigated the refolding and unfolding kinetics of the proteins by stopped-flow CD measurements. The results show that the rate of unfolding of the recombinant protein is faster than that of the authentic protein (Table 2), whereas the refolding rates are very similar in the two proteins (Table 3). This shows that the stability difference is caused by the enhanced unfolding rate of the recombinant protein, and this is interpreted in terms of the difference in the free energy of the unfolding transition ($\Delta\Delta G_U$) and the difference in the activation free energy ($\Delta\Delta G_U^{\ddagger}$) of unfolding. The $\Delta\Delta G_U^{\ddagger}$ is known to be given by the ratio of the unfolding rate constants as:

$$\Delta\Delta G_U^{\ddagger} = -RT \ln\left[\frac{k_u(\text{rec})}{k_u(\text{auth})}\right] \qquad (7)$$

where k_u(rec) and k_u(auth) represent the unfolding rate constants for the recombinant and authentic proteins, respectively. Because k_u(rec) is 5.7 times larger than k_u(auth) at 5.4 M GdnHCl, $\Delta\Delta G_U^{\ddagger}$ is estimated to be 1.0 kcal/mol, and this value is nearly identical with the $\Delta\Delta G_U$ (0.8 kcal/mol) at the same concentration of the denaturant. Thus, the stability difference between the proteins can be fully interpreted in terms of the increase in the unfolding rate of the recombinant protein. This means that the structure around the mutation site, the N terminus in this case, has not yet been organized in the transition state of refolding in α-lactalbumin (Kuwajima *et al.*, 1989; Matouschek *et al.*, 1989; Serrano *et al.*, 1992). The folding initiation site of α-lactalbumin is thus not located in the N-terminal region of the protein. Previous studies have shown that the structure around the 6-120 disulfide bond and that around the B helix, both of which are involved in the α-domain of this protein, have not yet been organized in the transition state of refolding (Ikeguchi *et al.*, 1998; T. Y. *et al.*, unpublished data), while the structure around the Ca^{2+}-binding site is known to be already organized in the transition state (Kuwajima *et al.*, 1989). Our results thus provide further support for the proposition that the folding initiation site of α-lactalbumin is located at the interface between the α and β-domains, around the Ca^{2+}-binding site of the protein.

Materials and Methods

Chemicals

GdnHCl was of a specially prepared reagent grade for biochemical use from Nacalai Tesque, Inc. (Kyoto). The concentration of GdnHCl was determined from the refractive index at 589 nm with an Atago 3T refractometer (Pace, 1986). Cyanogen bromide (CNBr) was purchased from Nacalai Tesque Inc. (Kyoto). Authentic goat α-lactalbumin was prepared from fresh goat milk by the method described (Kuwajima *et al.*, 1980). A Resource™-Q anion exchange column was purchased from Pharmacia Biotechnology, Inc. (Sweden) and a μ BONDASPHERE 5 μ C4 300 Å reversed-phase column was supplied by Nihon Waters Ltd (Japan).

Expression and purification of recombinant goat α-lactalbumin

The expression system of goat α-lactalbumin and the procedures for the refolding and purification of the protein have been reported by Kumagai *et al.* (1990) and recently improved by Uchiyama *et al.* (1995) utilizing a T7 promoter (Studier & Moffatt, 1986). In brief, the protein expressed in *E. coli* BL21(DE3) as inclusion bodies was solubilized in 8 M urea containing 20 mM Tris-HCl (pH 8.0) and first purified using a DEAE-Sepharose FF column. The eluted protein was reduced by 50 mM dithiothreitol and dialyzed against 20 mM Tris-HCl (pH 8.0) at 4 °C to remove urea. Refolding of the reduced α-lactalbumin was performed as described (Sawano *et al.*, 1992), with slight modifications, in a solution containing 20% (v/v) glycerol, 20 mM Tris-HCl (pH 8.0), 1 mM $CaCl_2$, 6 mM glutathione, 0.6 mM oxidized glutathione, 3.3 μM α-lactalbumin at 15 °C for more than 20 hours. The refolding process was monitored by the appearance of a sharp peak on a reversed-phase HPLC chromatogram detected by UV-absorbance at 215 nm using a C4 column with a linear gradient elution of 28%-52% acetonitrile in the presence of 0.1% (v/v) trifluoroacetic acid at a flow rate of 0.5 ml per minute. The refolded protein was then purified by DEAE-Sepharose FF and phenyl-Sepharose CL column chromatographies as described by Lindahl & Vogel (1984). Concentrations of authentic and recombinant goat α-lactalbumin were determined spectrophotometrically using an extinction coefficient of $E^{1\%}_{1\,cm} = 20.1$ for both (Kuwajima *et al.*, 1980). No free cysteinyl residues were detected in the folded recombinant protein by thiol content analysis (Ellman, 1959; Riddle *et al.*, 1979).

Preparation of methionine-free recombinant goat α-lactalbumin

The methionine-free protein was prepared according to the method described by Kim *et al.* (1997) with slight modifications. Recombinant goat α-lactalbumin was dissolved in 70% (v/v) formic acid and treated with 100 mM CNBr (50-100-fold molar excess over the protein concentration) for 24 hours in the dark at room temperature. The cleaved product was diluted ten times with water and dialyzed against 10 mM HCl, then dialyzed against 10 mM Tris-HCl (pH 8.5) containing 1 mM $CaCl_2$. Finally, the protein solution was purified on a Q-Sepharose FF column, which had been equilibrated with 20 mM Tris-HCl (pH 8.5) containing 1 mM $CaCl_2$ and eluted with a linear gradient of NaCl from 0 M to 0.5 M. The mobilities and retention times of the eluted

fractions were checked by native PAGE and anion-exchange HPLC, and compared with those of the authentic protein under the same conditions. The mass of the methionine-cleaved protein was determined by mass spectrometric analysis, and the removal of the N-terminal methionine residue was confirmed by the N-terminal sequence analysis. The concentration of the CNBr-cleaved protein was calculated using the same extinction coefficient as that given above.

Mass spectrometric analysis

Mass spectrometric analyses of the authentic, recombinant and methionine-free proteins were carried out by the MALDI-TOF-MS mass spectroscopic method. Sinapinic acid mix protein samples were used as the matrix, and the spectra were taken in Reflex (Bruker).

N-terminal sequence analysis

N-terminal sequencing of recombinant, authentic, and CNBr-cleaved proteins were carried out using an automated Applied Biosystem sequencer model 477a equipped with a model 120A on-line PTH amino acid analyzer. In this study we analyzed the first five residues in the proteins.

Equilibrium CD measurements

Equilibrium CD spectra were taken on a Jasco J-720 spectropolarimeter using an optical cuvette with a path length of 1.00 mm for measurements in the peptide region and 10.0 mm for measurements in the aromatic region. The CD spectra of the protein were measured in 50 mM sodium cacodylate, 50 mM NaCl (pH 7.0) containing 1 mM $CaCl_2$. The solutions for the GdnHCl-induced equilibrium unfolding studies were prepared in the same buffer containing various concentrations of GdnHCl. The mean residue ellipticity was calculated as a function of GdnHCl concentration at 25 °C by taking 113 as the mean residue mass. The protein concentration in the equilibrium measurements was 0.15-0.2 mg/ml.

The apparent fractional extent (F_{app}) of unfolding was calculated by:

$$F_{app} = \frac{\theta_{obs} - \theta_N}{\theta_U - \theta_N} \quad (8)$$

where θ_{obs} is the observed ellipticity, and θ_N and θ_U are the ellipticities in the native (N) and the fully unfolded (U) states, respectively. The θ_N and θ_U values are assumed to linearly depend on the GdnHCl concentration (C) as $\theta_N = \theta_1 + a_1 C$ and $\theta_U = \theta_2 + a_2 C$. The N state baseline was calculated from the ellipticity values between 0.5 and 2 M and between 0.4 and 1.8 M GdnHCl, and the U state baseline was from the values between 4.5 and 6.2 M and between 3.8 and 6.2 M GdnHCl for the authentic and recombinant proteins, respectively.

Kinetic measurements

Refolding and unfolding reactions of the authentic and recombinant proteins were induced by GdnHCl concentration jumps, which were performed by a stopped-flow CD apparatus (UNISOKU Inc., Japan) installed in the cell compartment of the J-720 spectropolarimeter sodium cacodylate at pH 7.0 and 25 °C. The dead time of the stopped-flow CD apparatus was 25 ms when a 4 mm cuvette was used. The concentration of the protein stock solution was about 1.5-2.0 mg/ml. The initial protein solutions before the concentration jump contained 1.0 M and 5.5 M GdnHCl for unfolding and refolding experiments, respectively. The diluent solution contained the same buffer (50 mM sodium cacodylate, 50 mM NaCl, and 1 mM $CaCl_2$, pH 7.0) and an appropriate concentration of GdnHCl. The two solutions were mixed with a mixing ratio of 1:10.

X-ray crystallographic studies

The crystal of recombinant goat α-lactalbumin was grown by the vapor diffusion method with a hanging drop in a chamber where the temperature was controlled at 20 °C. The data were collected by an automated area detector system, DIP2000, on an X-ray generator with a bent mirror system at 9.5 °C. Data processing and reduction was performed using DENZO and SCALEPACK programs (Otwinowski, 1993). The crystallographic data, the diffraction intensity statistics, and the refinement statistics are listed in Table 4. The crystal structure was solved on the basis of the model structure of baboon α-lactalbumin (Acharya *et al.*, 1989) by the molecular replacement method (Brünger, 1990) and was refined by a slow-cooling molecular-simulated annealing method in the X-PLOR 3.1 program suite (Brünger, 1992).

Theoretical estimation of $\Delta\Delta G_U$ between recombinant and authentic goat α-lactalbumin

The $\Delta\Delta G_U$ value was calculated by the method described by Oobatake & Ooi (1993). In this calculation, every atom was identified as belonging to one of seven atomic groups: aliphatic C, aromatic C, hydroxyl O, amide N, carbonyl C, carbonyl O, and sulphur S. In addition, the accessible surface area (ASA) of each atom in the N state (except hydrogen) was calculated by the method described by Richmond (1984) using the coordinates of the X-ray crystal structures. Because the N-terminal methionine residue is present only in the recombinant protein, the $\Delta\Delta G_U$ was assumed as a first approximation to be equal to the free energy change of unfolding (ΔG^u) of the methionine residue. For the ASA of atoms in the methionine residue in the U state, the values calculated by Shrake & Rupley (1973) were used. It was also assumed that the ΔG^u_h and ΔH^u_c are proportional to the change in the ASA ($\Delta\alpha_i$ for the ith atomic group) of the atoms on unfolding according to Oobatake & Ooi (1993). Thus:

$$\begin{aligned}
\Delta G^u &= \Delta G^u_h + \Delta G^u_c \\
\Delta G^u_h &= \Sigma_i g_{i,h} \Delta\alpha_i \\
\Delta G^u_c &= \Delta H^u_c - T\Delta S^u_c \\
\Delta H^u_c &= \Sigma_i h_{i,c} \Delta\alpha_i
\end{aligned} \quad (9)$$

where $g_{i,h}$ and $h_{i,c}$ are proportionality constants for the seven atomic groups. Although the change in the conformational entropy, ΔS^u_c, was also assumed to be proportional to the $\Delta\alpha_i$ values in the original Oobatake & Ooi (1993) method, this assumption may not be correct for the extra methionine residue of recombinant goat α-lactalbumin due to the rigid nature of this residue as

the distance between the C^{ε} atom of the methyl side-chain of Met0 and the C^{δ} atom of Gln2 side-chain being 3.5 Å. Moreover, the N-terminal amino group in the recombinant protein is hydrogen-bonded with the carbonyl oxygen atom of the Gln39 side-chain. Thus the $-T\Delta S_c^u$ value (−5.9 kcal/mol) obtained from Table 8 of Oobatake & Ooi (1993) was employed (see Discussion).

The contribution of other residues to the $\Delta\Delta G_U$ value was also estimated by $\Sigma_i\ (g_{i,h} + g_{i,c})\Delta\alpha_i^N$, where $g_{i,c}$ is a proportionality constant and $\Delta\alpha_i$ is the difference in the ASA value of the ith atomic group between the authentic and recombinant proteins in the N state. Here, $-T\Delta S_c^u$ was assumed to be proportional to the change in the ASA values following the original Oobatake & Ooi (1993) method. The values obtained for Mol A and Mol B of the crystal structure of the authentic protein were averaged. Since Glu1 of the authentic protein is more exposed to solvent in the unfolded state, the difference in the ASA of the atoms of Glu1 between the authentic and recombinant proteins in the unfolded state ($\Delta\alpha_i^U$) was also taken into account for the estimation of $\Delta\Delta G_U$ as $\Sigma_i\ (g_{i,h} + g_{i,c})\ \Delta\alpha_i^U$ using the ASA values of Shrake & Rupley (1973). The contribution of the other residues to the $\Delta\Delta G_U$ thus estimated has been found to be less than 1 kcal/mol.

Protein Data Bank accession number

The coordinates have been deposited in the Brookhaven Protein Data Bank with accession number 1HMK.

Acknowledgments

The authors thank Mr K. Maki, Mr T. Makio, Mr T. Hayashi, and Mr M. Mizuguchi for their assistance in this work. This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture and Science of Japan. T.K.C. is a postdoctoral Fellow of the Japan Society for the Promotion of Science.

References

Acharya, K. R., Stuart, D. I., Walker, N. P. C., Lewis, M. & Philips, D. C. (1989). Refined structure of baboon α-lactalbumin at 1.7 Å resolution. Comparison with C-type lysozyme. *J. Mol. Biol.* **208**, 99-127.

Acharya, K. R., Jingshan, R., Stuart, D. I., Philips, D. C. & Fenna, R. E. (1991). Crystal structure of human α-lactalbumin at 1.7 Å resolution. *J. Mol. Biol.* **221**, 571-581.

Adams, J. M. (1968). On the release of the formyl group from nascent protein. *J. Mol. Biol.* **33**, 571-589.

Arai, M. & Kuwajima, K. (1996). Rapid formation of a molten globule intermediate in refolding of α-lactalbumin. *Folding Des.* **1**, 275-287.

Aronsson, G., Martensson, L. G., Carlsson, U. & Jonsson, B. H. (1995). Folding and stability of the N-terminus of human carbonic anhydrase II. *Biochemistry*, **34**, 2153- 2162.

Balbach, J., Forge, V., Lau, W. S., van Nuland, N. A., Brew, K. & Dobson, C. M. (1996). Protein folding monitored at individual residues during a two-dimensional NMR experiment. *Science*, **274**, 1161-1163.

Brünger, A. T. (1990). Extension of molecular replacement: A new search strategy based on patterson correlation refinement. *Acta Crystallog. sect. A*, **46**, 46-57.

Brünger, A. T. (1992). *X-PLOR: Version 3.1. A system for crystallography and NMR*, Yale University Press, New Haven, CT.

Duverger, N., Murry-Brelier, A., Latta, M., Reboul, S., Castro, G., Mayaux, J. F., Fruchart, J. C., Taylor, J. M., Steinmetz, A. & Denefle, P. (1991). Functional characterization of human recombinant apolipoprotein AIV produced in *Escherichia coli*. *Eur. J. Biochem.* **201**, 373-383.

Ellman, G. L. (1959). Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **82**, 70-77.

Goldstein, J., Pollitt, N. S. & Inouye, M. (1990). Major cold shock protein of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **87**, 283-287.

Harata, K. & Muraki, M. (1992). X-ray structural evidence for a local helix-loop transition in α-lactalbumin. *J. Biol. Chem.* **267**, 1419-1421.

Hargrove, M. S., Krzywda, S., Wilkinson, A. J., Dou, Y., Ikeda-Saito, M. & Olson, J. S. (1994). Stability of myoglobin: a model for the folding of heme proteins. *Biochemistry*, **33**, 11767-11775.

Hiraoka, Y. & Sugai, S. (1985). Equilibrium and kinetic study of sodium- and potassium-induced conformational changes of apo-α-lactalbumin. *Int. J. Pept Protein Res.* **26**, 252-261.

Hynes, T. R. & Fox, R. O. (1991). The crystal structure of staphylococcal nuclease at 1.7 Å resolution. *Proteins: Struct. Funct. Genet.* **10**, 92-105.

Ikeguchi, M., Fujino, M., Kato, M., Kuwajima, K. & Sugai, S. (1998). Transition state in the folding of α-lactalbumin probed by the 6-120 disulfide bond. *Protein Sci.* **7**, 1564-1574.

Imoto, T., Yamada, H., Yasukochi, T., Yamada, E., Ito, Y., Ueda, T., Nagatani, H., Miki, T. & Horiuchi, T. (1987). Point mutation of alanine (31) to valine prohibits the folding of reduced lysozyme by sulfhydryl-disulfide interchange. *Protein Eng.* **4**, 333-338.

Ishikawa, N., Chiba, T., Chen, L. T., Shimizu, A., Ikeguchi, M. & Sugai, S. (1998). Remarkable destabilization of recombinant α-lactalbumin by an extraneous N-terminal methionyl residue. *Protein Eng.* **11**, 333-335.

Kataoka, M., Kuwajima, K., Tokunaga, F. & Goto, Y. (1997). Structural characterization of the molten globule of α-lactalbumin by solution X-ray scattering. *Protein Sci.* **6**, 422-430.

Katsumata, K., Okazaki, A., Tsurupa, G. P. & Kuwajima, K. (1996). Dominant forces in the recognition of a transient folding intermediate of α-lactalbumin by GroEL. *J. Mol. Biol.* **264**, 643-649.

Kim, S., Baum, J. & Anderson, S. (1997). Production of correctly folded recombinant [^{13}C, ^{15}N]-enriched guinea pig [Val90]- α-lactalbumin. *Protein Eng.* **10**, 455-462.

Kordel, J., Forsen, S. & Chazin, W. J. (1989). ^1H NMR sequential resonance assignments, secondary structure, and global fold in solution of the major (trans-Pro43) form of bovine calbindin D9k. *Biochemistry*, **28**, 7065-7074.

Kuhlman, B., Boice, J. A., Wu, W. J., Fairman, R. & Raleigh, D. P. (1997). Calcium binding peptides from α-lactalbumin: implications for protein folding and stability. *Biochemistry*, **36**, 4607-4615.

Kumagai, I., Takeda, S., Hibino, T. & Miura, K. (1990). Expression of goat α-lactalbumin in *Escherichia coli*

and its refolding to biologically active protein. *Protein Eng.* **3**, 449-452.

Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular protein structure. *Proteins: Struct. Funct. Genet.* **6**, 87-103.

Kuwajima, K. (1996). The molten globule state of α-lactalbumin. *FASEB J.* **10**, 102-109.

Kuwajima, K., Nitta, S. & Sugai, S. (1980). Intramolecular perturbation of tryptophans induced by the protonation of ionizable groups in goat α-lactalbumin. *Biochim. Biophys. Acta*, **623**, 389-401.

Kuwajima, K., Ogawa, Y. & Sugai, S. (1981). Role of the interaction between ionizable groups in the folding of bovine α-lactalbumin. *J. Biochem. (Tokyo)*, **89**, 759-770.

Kuwajima, K., Mitani, M. & Sugai, S. (1989). Characterization of the critical state in protein folding. Effects of guanidine hydrochloride and specific Ca^{2+} binding on the folding kinetics of α-lactalbumin. *J. Mol. Biol.* **206**, 547-561.

Lindahl, L. & Vogel, H. J. (1984). Metal-ion-dependent hydrophobic-interaction chromatography of α-lactalbumins. *Anal. Biochem.* **140**, 394-402.

Luzzati, P. V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallog.* **5**, 802-810.

Marcker, K. & Sanger, F. (1964). *N*-Formyl-methionyl-*S*-RNA. *J. Mol. Biol.* **8**, 354-360.

Matouschek, A., Kellis, J. T., Jr, Serrano, L. & Fersht, A. R. (1989). Mapping the transition state and pathway of protein folding by protein engineering. *Nature*, **340**, 122-126.

Miki, T., Yasukochi, Y., Nagatani, H., Fruno, M., Orita, T., Yamada, H., Imoto, T. & Horiuchi, T. (1987). Construction of a plasmid vector for the regulatable high level expression of eukaryotic genes in *Escherichia coli*: an application to over production of chicken lysozyme. *Protein Eng.* **4**, 327-332.

Milla, M. E., Brown, B. M. & Sauer, R. T. (1993). P22 Arc repressor: enhanced expression of unstable mutants by addition of polar C-terminal sequences. *Protein Sci.* **2**, 2198-2205.

Milla, M. E., Brown, B. M., Waldburger, C. D. & Sauer, R. T. (1995). P22 Arc repressor: transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry*, **343**, 13914-13919.

Miller, C. G., Strauch, K. L., Kurkal, A. M., Miller, J. L., Wingfield, P. T., Mazzei, G. J., Werlen, R. C., Graber, P. & Movva, N. R. (1987). N-terminal methionine-specific peptidase in *Salmonella typhimurium*. *Proc. Natl Acad. Sci. USA*, **84**, 2718-2722.

Mine, S., Ueda, T., Hashimoto, Y. & Imoto, T. (1997). Improvement of the refolding yield and solubility of hen egg-white lysozyme by altering the Met residue attached to its N-terminus to Ser. *Protein Eng.* **10**, 1333-1338.

Moerschell, R. P., Hosokawa, Y., Tsunasawa, S. & Sherman, F. (1990). The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine *in vivo*. Processing of altered iso-1-cytochromes *c* created by oligonucleotide transformation. *J. Biol. Chem.* **265**, 19638-19643.

Oobatake, M. & Ooi, T. (1993). Hydration and heat stab-

Otwinowski, Z. (1993). *DENZO: An Oscillation Data Processing for Macromolecular Crystallography*, Yale University, New Haven, C.T.

Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266-280.

Pfeil, W. (1998). Is the molten globule a third thermodynamic state of protein? The example of α-lactalbumin. *Proteins: Struct. Funct. Genet.* **30**, 43-48.

Pike, A. C., Brew, K. & Acharya, K. R. (1996). Crystal structures of guinea-pig, goat and bovine α-lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure*, **4**, 691-703.

Ptitsyn, O. B. (1995). Molten globule and protein folding. *Advan. Protein Chem.* **47**, 83-229.

Raumann, B. E., Rould, B. A., Pabo, C. O. & Sauer, R. T. (1994). DNA recognition by β-sheets in the Arc repressor-operator crystal structure. *Nature*, **367**, 754-757.

Reid, K. L., Rodriguez, H. M., Hillier, B. J. & Gregoret, L. M. (1998). Stability and folding properties of a model β-sheet protein, *Escherichia coli* CspA. *Protein Sci.* **7**, 470-479.

Richmond, T. L. (1984). Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* **178**, 63-89.

Riddle, P. W., Blakeley, R. L. & Zerner, B. (1979). Ellman's reagent: 5,5'-dithiobis (2-nitrobenzoic acid)-a reexamination. *Anal. Biochem.* **94**, 75-81.

Sawano, H., Koumoto, Y., Ohta, K., Sasaki, Y., Segawa, S. & Tachibana, H. (1992). Efficient in vitro folding of the three-disulfide derivatives of hen lysozyme in the presence of glycerol. *FEBS Letters*, **303**, 11-14.

Schulman, B. A. & Kim, P. S. (1996). Proline scanning mutagenesis of a molten globule reveals non-cooperative formation of a protein's overall topology. *Nature Struct. Biol.* **3**, 682-687.

Schulman, B. A., Kim, P. S., Dobson, C. M. & Redfield, C. (1997). A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nature Struct. Biol.* **4**, 630-634.

Schultz, D. A. & Baldwin, R. L. (1992). *Cis* proline mutants of ribonuclease A. I. Thermal stability. *Protein Sci.* **1**, 910-916.

Serrano, L., Matouschek, A. & Fersht, A. R. (1992). The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 805-818.

Shrake, A. & Rupley, J. A. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351-371.

Shimizu, A., Ikeguchi, M., Kobayashi, T. & Sugai, S. (1996). A synthetic peptide study on the molten globule of α-lactalbumin. *J. Biochem. (Tokyo)*, **119**, 947-952.

Shortle, D. (1996). The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* **10**, 27-34.

Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130.

Suciu, D. & Inouye, M. (1996). The 19-residue pro-peptide of staphylococcal nuclease has a profound

Sugai, S. & Ikeguchi, M. (1994). Conformational comparison between α-lactalbumin and lysozyme. *Advan. Biophys.* **30**, 37-84.

Takeda, M. & Webster, R. (1968). Protein chain initiation and deformylation in *B. subtilis* homogenates. *Proc. Natl Acad. Sci. USA*, **60**, 1487-1494.

Tanford, C. (1970). Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Advan. Protein Chem.* **245**, 4760-4769.

Uchiyama, H., Perez-Prat, E. M., Watanabe, K., Kumagai, I. & Kuwajima, K. (1995). Effects of amino acid substitutions in the hydrophobic core of α-lactalbumin on the stability of the molten globule state. *Protein Eng.* **8**, 1153-1161.

Vanderheeren, G. & Hanssens, I. (1994). Thermal unfolding of bovine α-lactalbumin. Comparison of circular dichroism with hydrophobicity measurements. *J. Biol. Chem.* **269**, 7090-7094.

Wilson, G., Hecht, L. & Barron, L. D. (1996). The native-like tertiary fold in molten globule α-lactalbumin appears to be controlled by a continuous phase transition. *J. Mol. Biol.* **261**, 341-347.

Wu, L. C. & Kim, P. S. (1997). Hydrophobic sequence minimization of the α-lactalbumin molten globule. *Proc. Natl Acad. Sci. USA*, **94**, 14314-14319.

Edited by P. E. Wright

(Received 22 June 1998; received in revised form 20 October 1998; accepted 22 October 1998)